

computation social science

Experiments & interventions at scale

Course structure

Period IV

Week	Lecture	Exer. dl	Ext. dl	Topic
1	Feb 27	Mar 3	Mar 15	Introduction to CSS
2	Mar 6	Mar 10	Mar 22	Artificial societies & agent-based models
3	Mar 13	Mar 17	Mar 29	Data & digital traces
4	Mar 20	Mar 24	Apr 5	Counting things & analysing text
5	Mar 27	Mar 31	Apr 12	Social networks: structure
6	Apr 3	*	-	Introduction to the project

Period V

Week	Lecture	Exercise dl	Ext. dl	Topic
7	Apr 24	May 5	May 10	Ethics, privacy, legal
-	-	-	-	WAPPU
8	May 8	May 12**	May 24	Agent-based models & emergence
9	May 15	May 19***	May 31	Social networks: dynamics
10	May 22	May 26***	June 7	Experiments & interventions at scale
11	May 29	-	-	Computing for social good

*Project deadline: May 26

Project peer review: June 2

**Bonus round

***Only lecture questions

Peer reviews

- Link to peer grading system on Mycourses: <https://mycourses.aalto.fi/mod/lti/view.php?id=1043129>
- Each person reviews 2 reports (35 points each)
- Two parts to the reports -
 - Task A reviews (10 points per report)
 - Task C or D reviews (25 points per report) – if the report include both, only review the first
- Deadline - 2nd June (at 19:00)

Reviews for Task A

s1

Reviews for Task A (Reviewer points: 20 points total, 10 per review)

Q1 · Scale

How well were the **primary goals** explained in the report?

Response

- Explanation could have been improved
- Sufficiently well explained
- Excellent explanation

Q2 · Text · Optional question

Any comments about the **primary goals**?

Response

Write your response here

Reviews for Task A

s1

Reviews for Task A (Reviewer points: 20 points total, 10 per review)

Q1 · Scale

How well were the **primary goals** explained in the report?

Response

- Explanation could have been improved
- Sufficiently well explained
- Excellent explanation

Q2 · Text · Optional question

Any comments about the **primary goals**?

Response

Write your response here

Q3 · Scale

How well were the **main methods** explained in the report?

Response

- Explanation could have been improved
- Sufficiently well explained
- Excellent explanation

Q4 · Text · Optional question

Any comments about the **main methods**?

Response

Write your response here

Q5 · Scale

How well were the **key findings** explained in the report?

Response

- Explanation could have been improved
- Sufficiently well explained
- Excellent explanation

Q6 · Text · Optional question

Any comments about the **key findings**?

Response

Write your response here

Reviews for Task A

s1

Reviews for Task A (Reviewer points: 20 points total, 10 per review)

Q1 · Scale

How well were the **primary goals** explained in the report?

Response

- Explanation could have been improved
- Sufficiently well explained
- Excellent explanation

Q2 · Text · Optional question

Any comments about the **primary goals**?

Response

Write your response here

Q3 · Scale

How well were the **main methods** explained in the report?

Response

- Explanation could have been improved
- Sufficiently well explained
- Excellent explanation

Q4 · Text · Optional question

Any comments about the **main methods**?

Response

Write your response here

Q7 · Text

Comment on the **follow up analysis**: Did you understand the idea that was introduced? How interesting was it?

Response

Write your response here

Q5 · Scale

How well were the **key findings** explained in the report?

Response

- Explanation could have been improved
- Sufficiently well explained
- Excellent explanation

Q6 · Text · Optional question

Any comments about the **key findings**?

Response

Write your response here

Reviews for Task C or D

S2

Review for Task C or D (Reviewer points: 50 points total, 25 per review)

Explanation

Here the report should include either Task C or Task D, and you might have completed only one of these tasks. That is okay, and the review questions can be answered without completing the tasks.

If the report has both tasks completed for some reason, please just review whichever comes first in the report.

Q8 · Yes / No

Did you complete the same task as the report (C or D)?

Response

No

Yes

Reviews for Task C or D

S2

Review for Task C or D (Reviewer points: 50 points total, 25 per review)

Explanation

Here the report should include either Task C or Task D, and you might have completed only one of these tasks. That is okay, and the review questions can be answered without completing the tasks.

If the report has both tasks completed for some reason, please just review whichever comes first in the report.

Q8 · Yes / No

Did you complete the same task as the report (C or D)?

Response

No

Yes

Q9 · Text

The report should have either Task C or D described in it. Depending on which one the student had chosen answer the following questions:

Task C

1. Did you understand the problem statement?
2. Was the method that was chosen appropriate to answer the problem?
3. Were you convinced of the results?
4. How interesting was the problem statement (and the results)?

Task D

1. *Part 1:* Comment on the clarity and how convincing was the answer to: Critical examination of analysis of the dataset for obtaining the results, with respect to the big data characteristics?
2. In *Part 2:* Did you understand primary goal, the main methods, and key findings of the reviewed paper solely based on the description in the report? (Given that you might not have read the paper, and if you have not, you don't need to read it).
3. In *Part 2:* Comment on the follow up analysis: Did you understand the idea that was introduced? How interesting was it?
4. In *Part 2:* Comment on the clarity and how convincing was the answer to: Critical examination of analysis of the dataset for obtaining the results, with respect to the big data characteristics?

Response

Write your response here

Reviews for Task C or D

S2

Review for Task C or D (Reviewer points: 50 points total, 25 per review)

Explanation

Here the report should include either Task C or Task D, and you might have completed only one of these tasks. That is okay, and the review questions can be answered without completing the tasks.

If the report has both tasks completed for some reason, please just review whichever comes first in the report.

Q8 · Yes / No

Did you complete the same task as the report (C or D)?

Response

No

Yes

Q9 · Text

The report should have either Task C or D described in it. Depending on which one the student had chosen answer the following questions:

Task C

1. Did you understand the problem statement?
2. Was the method that was chosen appropriate to answer the problem?
3. Were you convinced of the results?
4. How interesting was the problem statement (and the results)?

Task D

1. *Part 1:* Comment on the clarity and how convincing was the answer to: Critical examination of analysis of the dataset for obtaining the results, with respect to the big data characteristics?
2. In *Part 2:* Did you understand primary goal, the main methods, and key findings of the reviewed paper solely based on the description in the report? (Given that you might not have read the paper, and if you have not, you don't need to read it).
3. In *Part 2:* Comment on the follow up analysis: Did you understand the idea that was introduced? How interesting was it?
4. In *Part 2:* Comment on the clarity and how convincing was the answer to: Critical examination of analysis of the dataset for obtaining the results, with respect to the big data characteristics?

Response

Write your response here

Q10 · Scale

How good was the overall quality of the report/results for this task?

Response

- Could have been improved
- Sufficiently quality
- Excellent quality

Course structure

Period IV

Week	Lecture	Exer. dl	Ext. dl	Topic
1	Feb 27	Mar 3	Mar 15	Introduction to CSS
2	Mar 6	Mar 10	Mar 22	Artificial societies & agent-based models
3	Mar 13	Mar 17	Mar 29	Data & digital traces
4	Mar 20	Mar 24	Apr 5	Counting things & analysing text
5	Mar 27	Mar 31	Apr 12	Social networks: structure
6	Apr 3	*	-	Introduction to the project

Period V

Week	Lecture	Exercise dl	Ext. dl	Topic
7	Apr 24	May 5	May 10	Ethics, privacy, legal
-	-	-	-	WAPPU
8	May 8	May 12**	May 24	Agent-based models & emergence
9	May 15	May 19***	May 31	Social networks: dynamics
10	May 22	May 26***	June 7	Experiments & interventions at scale
11	May 29	-	-	Computing for social good

*Project deadline: May 26

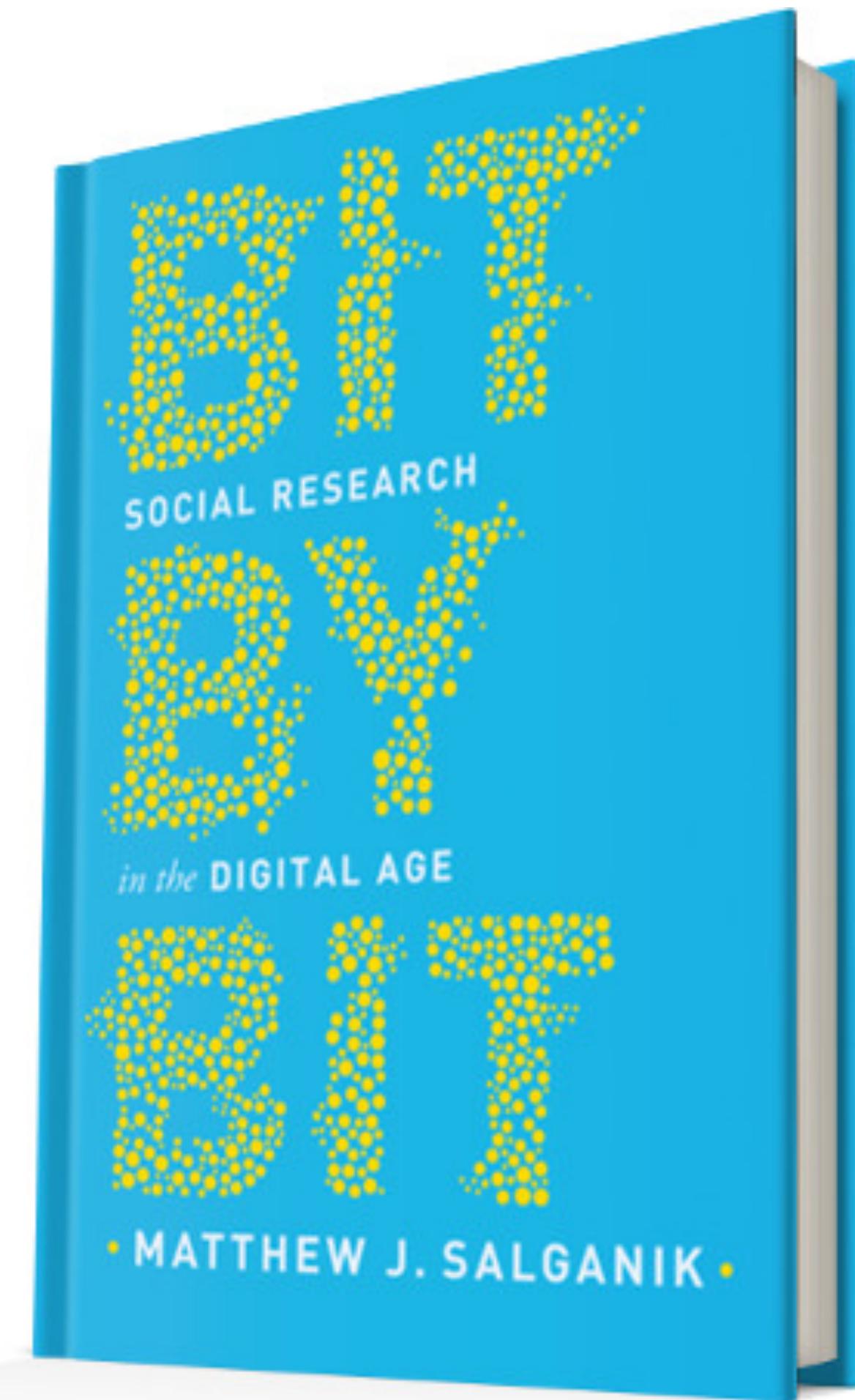
Project peer review: June 2

**Bonus round

***Only lecture questions

Running experiments

Chapter 4



<https://www.bitbybitbook.com/en/1st-ed/>

Experimental Study of Informal Rewards in Peer Production

Michael Restivo*, Arnout van de Rijt

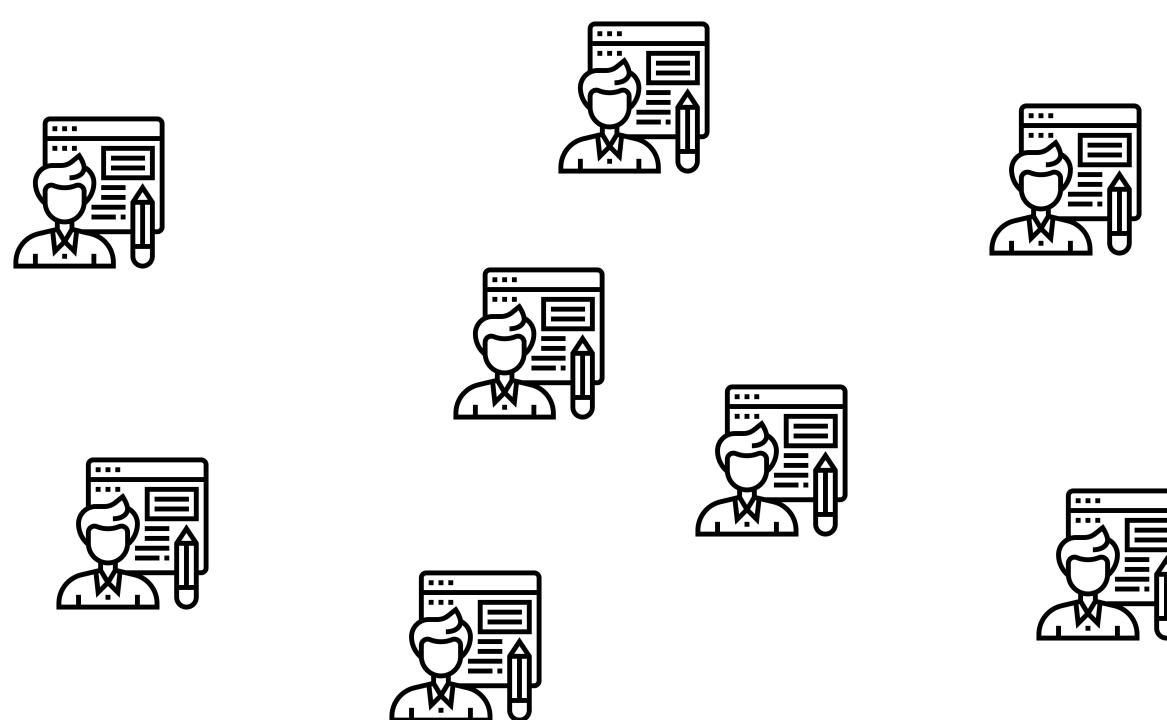
Department of Sociology, State University of New York at Stony Brook, New York, New York, United States of America

Abstract

We test the effects of informal rewards in online peer production. Using a randomized, experimental design, we assigned editing awards or “barnstars” to a subset of the 1% most productive Wikipedia contributors. Comparison with the control group shows that receiving a barnstar increases productivity by 60% and makes contributors six times more likely to receive additional barnstars from other community members, revealing that informal rewards significantly impact individual effort.

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0034358>

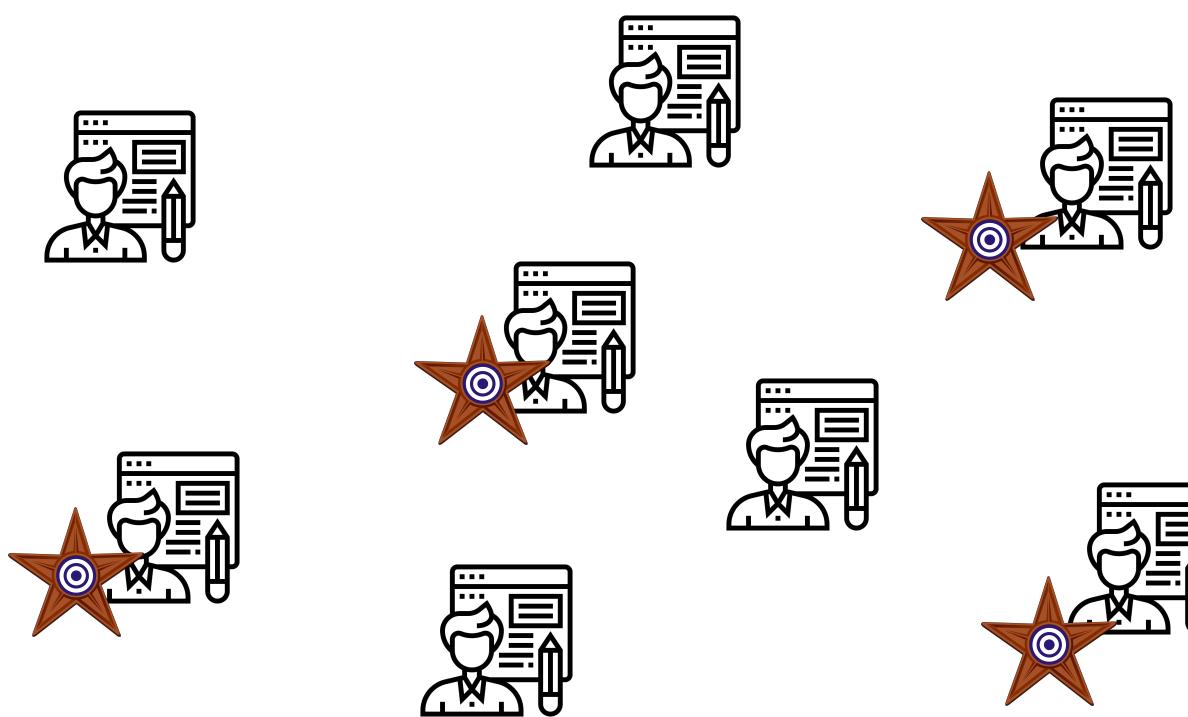
Effect of awards on contributions to Wikipedia



- Why do people contribute to public goods where everyone can benefit from them and not just the people who contributed to it?
- Long been of interest to social scientists
- Wikipedia is a good example of public goods where contributions are recorded, so easy to study



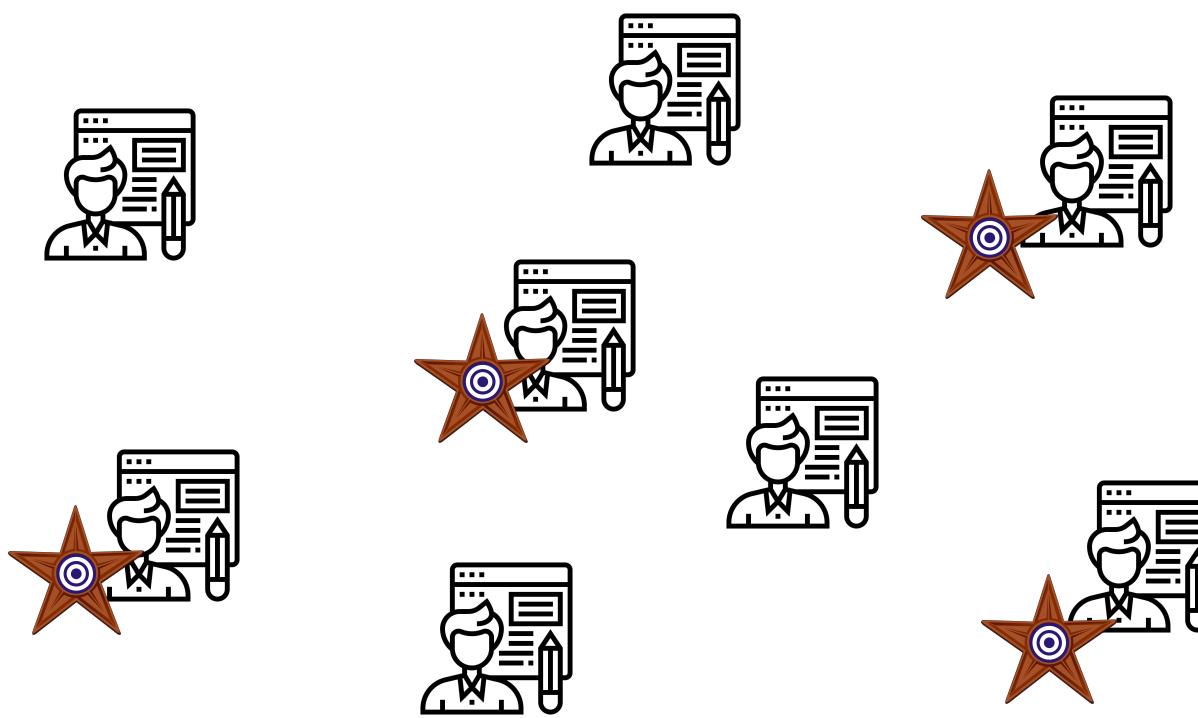
Effect of awards on contributions to Wikipedia



- Give editors awards - barn stars - sign of recognition in Wikipedia editors' community
- Selected deserving (top 1% of contributors) editors who hadn't received barn stars yet and gave a barn star to a random subset of them
- Examine how much do these editors edit after receiving barn stars

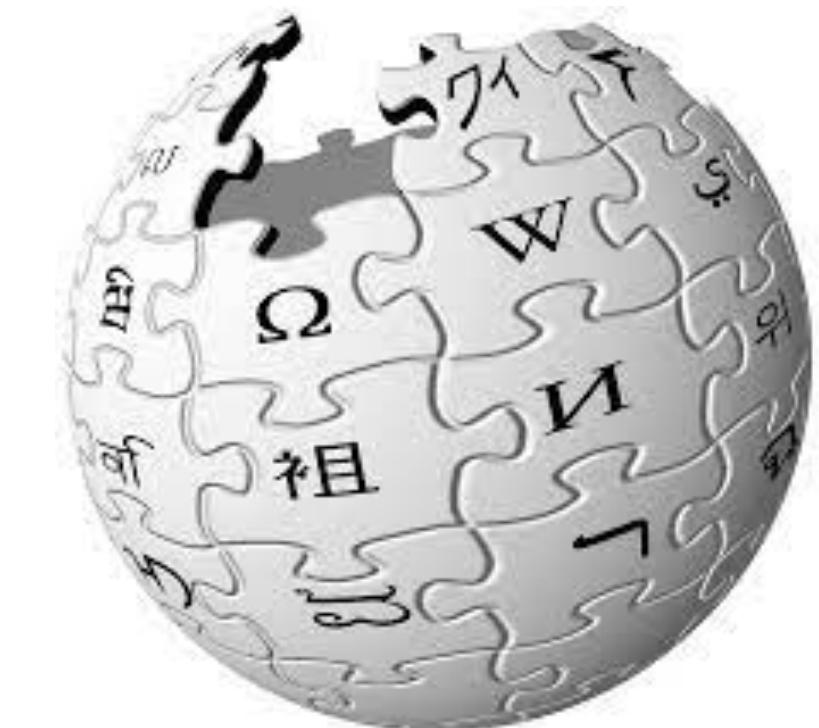


Effect of awards on contributions to Wikipedia

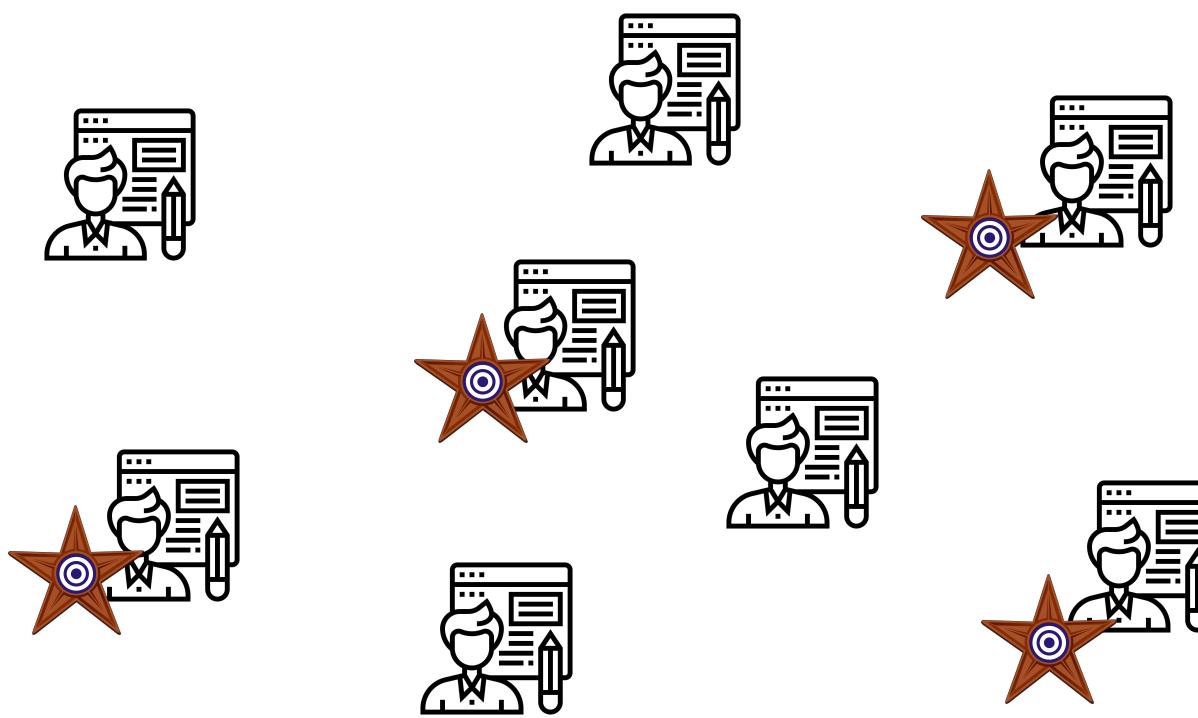


Treatment group - Editors who received barn stars edited less afterwards

- Contrary to the intuition that getting awards might encourage you to contribute more



Effect of awards on contributions to Wikipedia



Treatment group - Editors who received barn stars edited less afterwards

- Contrary to the intuition that getting awards might encourage you to contribute more

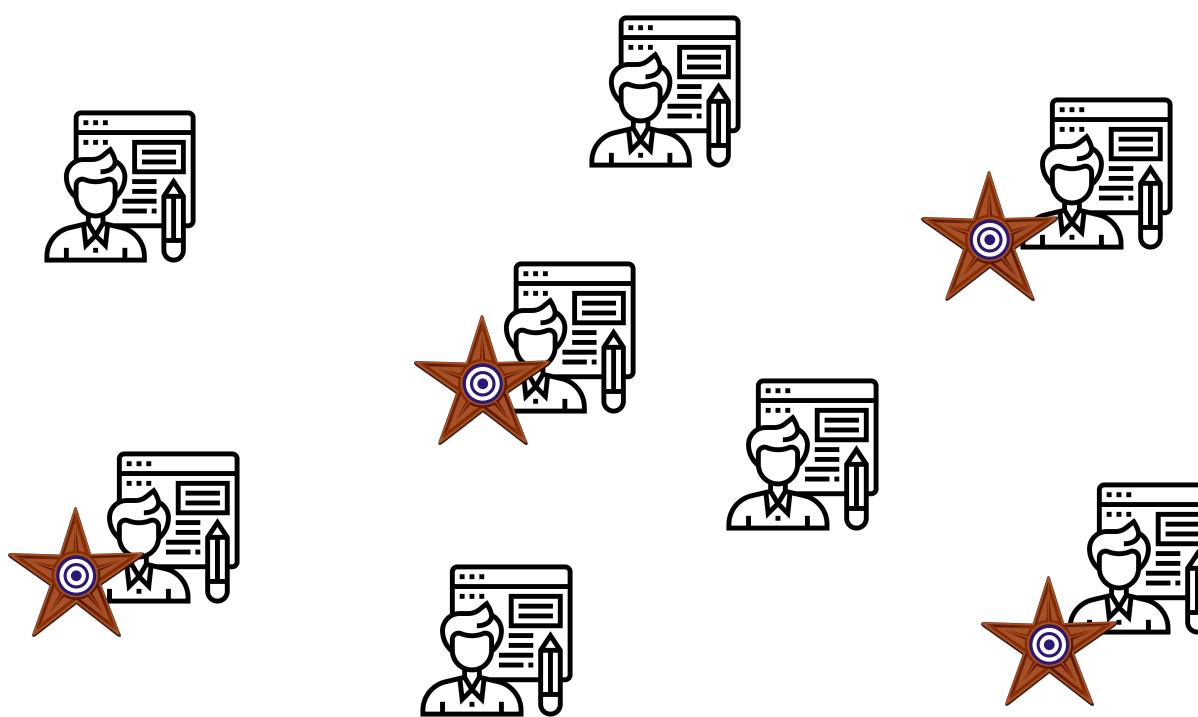


Control group - deserving editors who didn't receive an award

- Contributed even lesser



Effect of awards on contributions to Wikipedia



Treatment group - Editors who received barn stars edited less afterwards

- Contrary to the intuition that getting awards might encourage you to contribute more



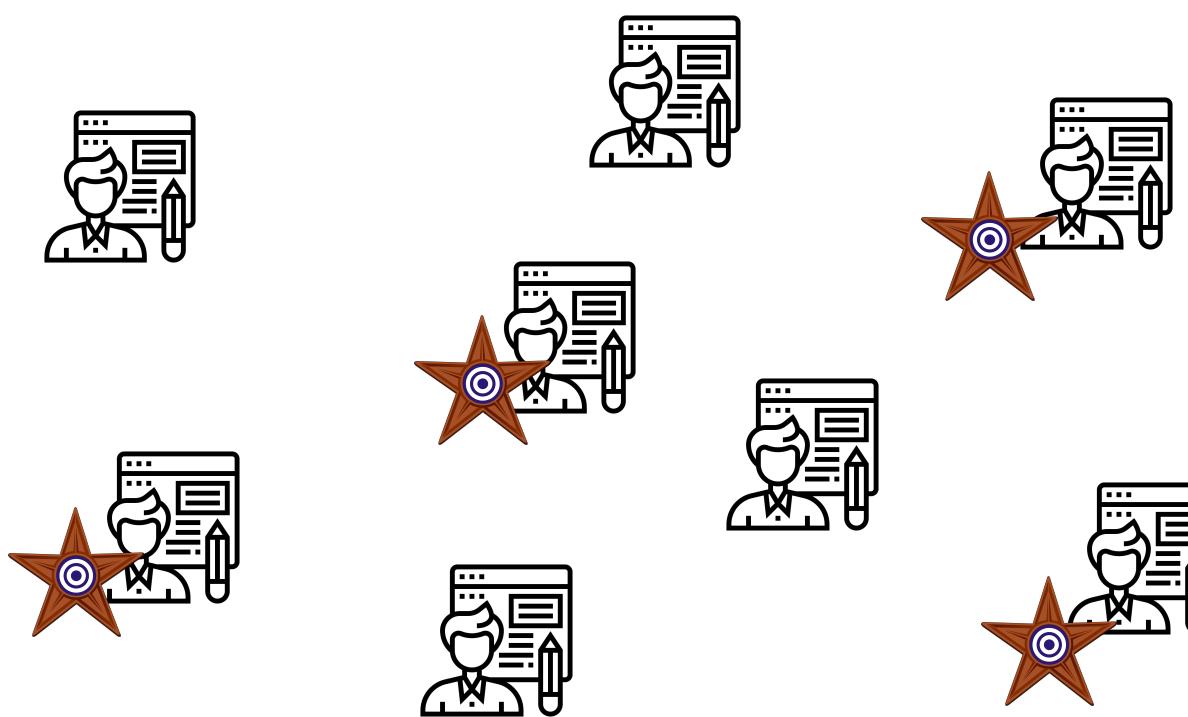
Control group - deserving editors who didn't receive an award

- Contributed even lesser

Awards increased the contributions in comparison to what would have happened without them



Effect of awards on contributions to Wikipedia



Bursty editing behavior on
Wikipedia!!



Treatment group - Editors who received barn stars edited less afterwards

- Contrary to the intuition that getting awards might encourage you to contribute more



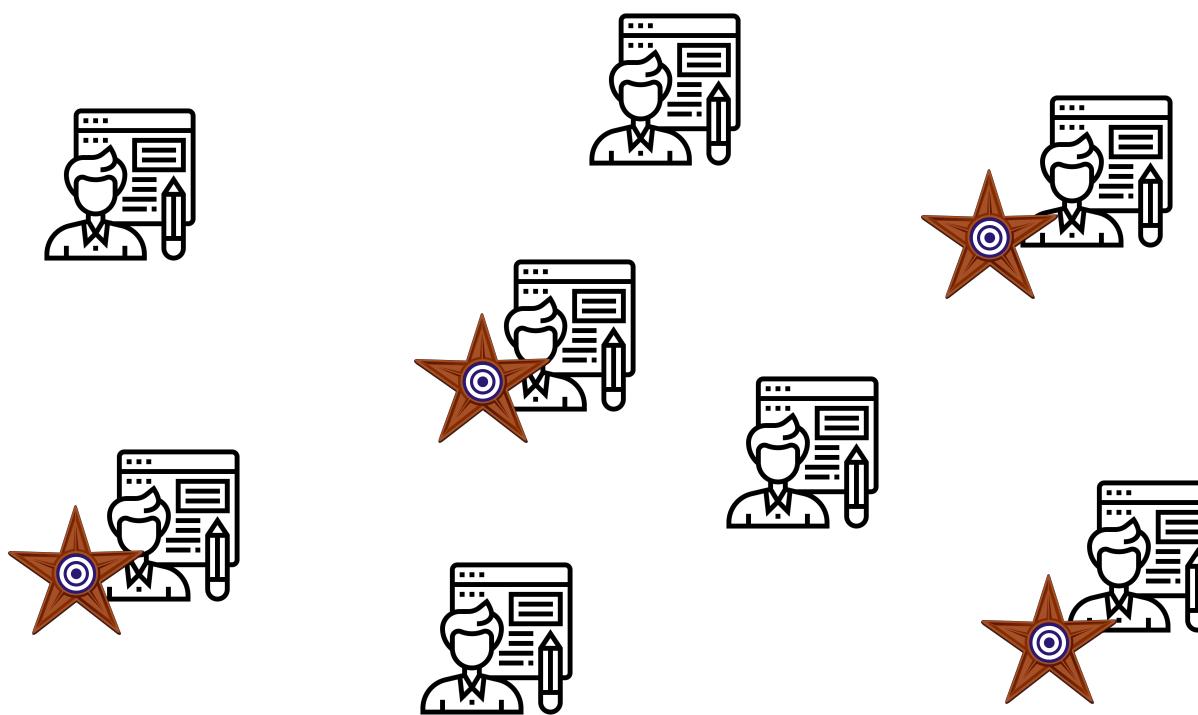
Control group - deserving editors who didn't receive an award

- Contributed even lesser

Awards increased the contributions in comparison to what would have happened without them



Effect of awards on contributions to Wikipedia



Without control group would have reached exactly the wrong conclusion



Treatment group - Editors who received barn stars edited less afterwards

- Contrary to the intuition that getting awards might encourage you to contribute more



Control group - deserving editors who didn't receive an award

- Contributed even lesser



Effect of awards on contributions to Wikipedia

Without control group would have reached exactly the wrong conclusion

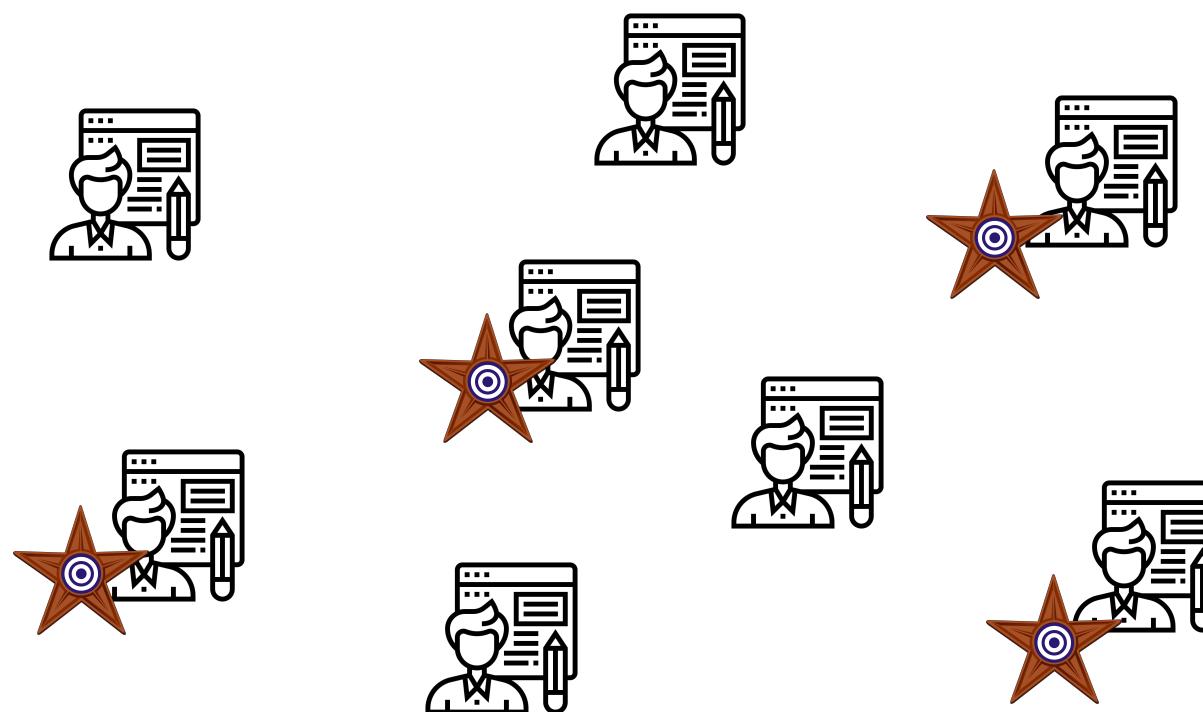


Treatment group - Editors who received barn stars edited less afterwards



Control group - deserving editors who didn't receive an award

- Contributed even lesser



Illustrates power of **randomised controlled experiments** - compare outcomes for treatment and control groups to estimate the effect of the treatment

Awards increased the contributions in comparison to what would have happened without them



Randomised controlled experiments

vs.

Perturb and observe experiments

What are experiments

- Recruit participants
- Randomised treatment & control
- Delivering treatment
- Measuring outcomes

What are experiments

- Recruit participants
- Randomised treatment & control
- Delivering treatment
- Measuring outcomes

OPEN  ACCESS Freely available online



Experimental Study of Informal Rewards in Peer Production

Michael Restivo*, Arnout van de Rijt

Department of Sociology, State University of New York at Stony Brook, New York, New York, United States of America

Abstract

We test the effects of informal rewards in online peer production. Using a randomized, experimental design, we assigned editing awards or “barnstars” to a subset of the 1% most productive Wikipedia contributors. Comparison with the control group shows that receiving a barnstar increases productivity by 60% and makes contributors six times more likely to receive additional barnstars from other community members, revealing that informal rewards significantly impact individual effort.

What are experiments

- Recruit participants
- Randomised treatment & control
- Delivering treatment
- Measuring outcomes
 - Fully digital experiment
 - Could they have increased the size of the experiment easily?

OPEN  ACCESS Freely available online

 PLOS one

Experimental Study of Informal Rewards in Peer Production

Michael Restivo*, Arnout van de Rijt

Department of Sociology, State University of New York at Stony Brook, New York, New York, United States of America

Abstract

We test the effects of informal rewards in online peer production. Using a randomized, experimental design, we assigned editing awards or “barnstars” to a subset of the 1% most productive Wikipedia contributors. Comparison with the control group shows that receiving a barnstar increases productivity by 60% and makes contributors six times more likely to receive additional barnstars from other community members, revealing that informal rewards significantly impact individual effort.

What are experiments

- Recruit participants
- Randomised treatment & control
- Delivering treatment
- Measuring outcomes
 - Fully digital experiment
 - Could they have increased the size of the experiment easily?
 - Yes, zero variable cost data, i.e., no cost to adding more participants
 - But ethical considerations — no harm to individual editors but harm to the wikipedia editing community

OPEN  ACCESS Freely available online

 PLOS one

Experimental Study of Informal Rewards in Peer Production

Michael Restivo*, Arnout van de Rijt

Department of Sociology, State University of New York at Stony Brook, New York, New York, United States of America

Abstract

We test the effects of informal rewards in online peer production. Using a randomized, experimental design, we assigned editing awards or “barnstars” to a subset of the 1% most productive Wikipedia contributors. Comparison with the control group shows that receiving a barnstar increases productivity by 60% and makes contributors six times more likely to receive additional barnstars from other community members, revealing that informal rewards significantly impact individual effort.

Dimensions of experiments

Common technique in social sciences

Lab vs. Field experiments

Continuum, but we focus on the extremes



Dimensions of experiments

What do you understand from lab experiments?



Dimensions of experiments

- Artificial setting - lab on campus
- Artificial task - does not typically occur in a normal environment
- Artificial stakes - people paid to do tasks, unrelated to what they do in experiment
- Participants not from target population of interest - convenience sample of college students
- Great for testing subtle theories which require very precise control of the environment



Dimensions of experiments

What do you understand from field experiments?



Dimensions of experiments

- Artificial setting - lab on campus
- Artificial task - does not typically occur in a normal environment
- Artificial stakes - people paid to do tasks, unrelated to what they do in experiment
- Participants not from target population of interest - convenient sample of college students
- Great for testing subtle theories which require very precise control of the environment
- Experiments in the wild
- Tasks or treatments are more likely to occur in natural environment
- Higher stakes - real outcomes in participants' lives
- Participants more similar to target population of interest that would normally receive the treatment
- Considered more realistic, but lesser control in field => harder to do more subtle manipulations

Lab experiments ← → **Field experiments**

Dimensions of experiments

JOURNAL ARTICLE

Getting a Job: Is There a Motherhood Penalty?

Shelley J. Correll, Stephen Benard, In Paik

American Journal of Sociology, Vol. 112, No. 5 (March 2007), pp. 1297-1338 (42 pages)

<https://www.jstor.org/stable/10.1086/511799> • <https://doi.org/10.1086/511799>

- Sources of “motherhood penalty”
- Mothers earn less money than childless women with similar skills and jobs
- Run both lab and field experiment



Dimensions of experiments

JOURNAL ARTICLE

Getting a Job: Is There a Motherhood Penalty?

Shelley J. Correll, Stephen Benard, In Paik

American Journal of Sociology, Vol. 112, No. 5 (March 2007), pp. 1297-1338 (42 pages)

<https://www.jstor.org/stable/10.1086/511799> • <https://doi.org/10.1086/511799>

- Sources of “motherhood penalty”
- Mothers earn less money than childless women with similar skills and jobs
- Run both lab and field experiment

Illustrates why instead of competing methods, it is better to think of lab and field experiments as complementary methods with different strengths and weaknesses.



Dimensions of experiments

JOURNAL ARTICLE

Getting a Job: Is There a Motherhood Penalty?

Shelley J. Correll, Stephen Benard, In Paik

American Journal of Sociology, Vol. 112, No. 5 (March 2007), pp. 1297-1338 (42 pages)

<https://www.jstor.org/stable/10.1086/511799> • <https://doi.org/10.1086/511799>

- Participants - college undergraduates
- Task - help with hiring for a lead position at a company
 - review resumes of potential candidates
 - rate candidates on intelligence, warmth and commitment to work
 - recommend whether to hire or not,
 - if yes, then what starting salary
- Same resumes, except treatment signalled motherhood, control didn't
- Less likely to hire mothers, with lower starting salary
- Statistical analysis revealed disadvantages for mothers largely explained by lower ratings in competence and commitment



Lab experiments



Field experiments

Dimensions of experiments

JOURNAL ARTICLE

Getting a Job: Is There a Motherhood Penalty?

Shelley J. Correll, Stephen Benard, In Paik

American Journal of Sociology, Vol. 112, No. 5 (March 2007), pp. 1297-1338 (42 pages)

<https://www.jstor.org/stable/10.1086/511799> • <https://doi.org/10.1086/511799>

- Participants - college undergraduates
- Task - help with hiring for a lead position at a company
 - review resumes of potential candidates
 - rate candidates on intelligence, warmth and commitment to work
 - recommend whether to hire or not,
 - if yes, then what starting salary
- Same resumes, except treatment signalled motherhood, control didn't
- Less likely to hire mothers, with lower starting salary
- Statistical analysis revealed disadvantages for mothers largely explained by lower ratings in competence and commitment
- Responded to hundred of job advertisements with fake cover letters and resumes, some singled motherhood, some didn't
- Participants - real companies
- Task - invited to interviews or not based on resume + cover letter
- Same resumes, except treatment signalled motherhood, control didn't
- Mothers were less likely to get called back for interviews than equally qualified childless women
- Don't know the reasons for the behavior, no process data available to understand the mechanisms behind decisions made by participants

Lab experiments



Field experiments

Dimensions of experiments

JOURNAL ARTICLE

Getting a Job: Is There a Motherhood Penalty?

Shelley J. Correll, Stephen Benard, In Paik

American Journal of Sociology, Vol. 112, No. 5 (March 2007), pp. 1297-1338 (42 pages)

- **Advantages of Lab over Field experiments**
 - Participants - college undergraduates
 - Task - help with hiring for a lab
 - review resumes of potential candidates
 - rate candidates on intelligence, competence, and commitment
 - recommend whether to hire or not
 - if yes, then what starting salary
 - Same resumes, except treatment
 - mothers vs. non-mothers
 - Less likely to hire mothers, with or without motherhood
 - Statistical analysis revealed disadvantages for mothers largely explained by lower ratings in competence and commitment
 - mothers were less likely to be hired than fathers
 - mothers were offered lower starting salaries than fathers
 - mothers were less likely to receive job offers than fathers
 - mothers were less likely to receive job interviews than equally qualified fathers
 - mothers were less likely to receive job offers than fathers
 - Don't know the reasons for the behavior, no process data available to understand the mechanisms behind decisions made by participants

Lab experiments



Field experiments

Dimensions of experiments

JOURNAL ARTICLE

Getting a Job: Is There a Motherhood Penalty?

Shelley J. Correll, Stephen Benard, In Paik

American Journal of Sociology, Vol. 112, No. 5 (March 2007), pp. 1297-1338 (42 pages)

- **Advantages of Lab over Field experiments**
 - Near total control of environment - resumes read in quite lab setting, some resumes may not even have been read in field
 - Additional process data to explain why participants are making their decisions in lab, not available for field settings
 - Participants - real companies
- Participants - college undergraduates
 - review resumes of potential candidates
 - rate candidates on intelligence, warmth and commitment to work
 - recommend whether to hire or not
 - if yes, then what starting salary
- Task - help with hiring for a lead position
 - resume + cover letter
- Same resumes, except treatment
 - motherhood, control didn't
- Less likely to hire mothers, without
 - r interviews than equally
- Statistical analysis revealed disadvantages for mothers largely explained by lower ratings in competence and commitment
 - Don't know the reasons for the behavior, no process data available to understand the mechanisms behind decisions made by participants

Lab experiments



Field experiments

Dimensions of experiments

JOURNAL ARTICLE

Getting a Job: Is There a Motherhood Penalty?

Shelley J. Correll, Stephen Benard, In Paik

American Journal of Sociology, Vol. 112, No. 5 (March 2007), pp. 1297-1338 (42 pages)

- **Advantages of Lab over Field experiments**
 - Near total control of environment - resumes read in quite lab setting, some resumes may not even have been read in field
 - Additional process data to explain why participants are making their decisions in lab, not available for field settings
 - Participants - real companies
- Participants - college undergraduates
 - Participants - real companies
- Task - help with hiring for a lead position
 - review resumes of potential candidates
 - rate candidates on intelligence, warmth and commitment to work
 - recommend whether to hire or not.
 - if yes, then what starting salary
- Disadvantages of Lab over Field experiments
 - Took - invited to interviews or not based on resume + cover letter
- Same resumes, except treatment
 - Took - invited to interviews or not based on resume + cover letter
- Less likely to hire mothers, with
 - Took - invited to interviews or not based on resume + cover letter
- Statistical analysis revealed disadvantages for mothers largely explained by lower ratings in competence and commitment
 - Don't know the reasons for the behavior, no process data available to understand the mechanisms behind decisions made by participants

Lab experiments



Field experiments

Dimensions of experiments

JOURNAL ARTICLE

Getting a Job: Is There a Motherhood Penalty?

Shelley J. Correll, Stephen Benard, In Paik

American Journal of Sociology, Vol. 112, No. 5 (March 2007), pp. 1297-1338 (42 pages)

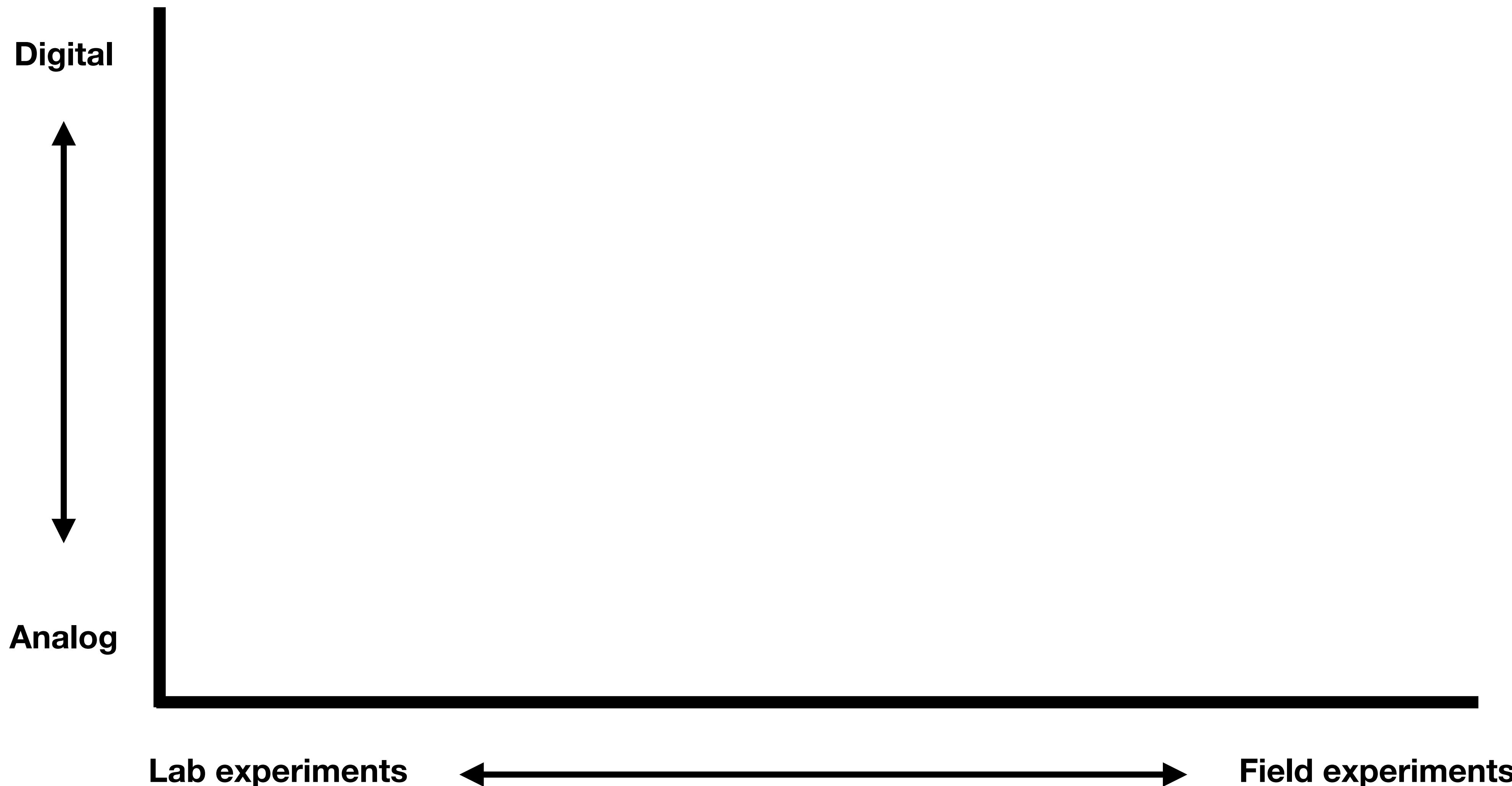
- **Advantages of Lab over Field experiments**
 - Near total control of environment - resumes read in quite lab setting, some resumes may not even have been read in field
 - Additional process data to explain why participants are making their decisions in lab, not available for field settings
 - Participants - real companies
- Participants - college undergraduates
 - Participants in lab could have guessed the goal of study and modified their behavior to not appear biased
 - Small differences in resumes can only stand out in very clean sterile lab environment - leading to overestimation of effect
 - Mothers were less likely to get called back for interviews than equally qualified childless women
 - Unreal participants
 - Don't know the reasons for the behavior, no process data available to understand the mechanisms behind decisions made by participants
- Task - help with hiring for a lead position
 - review resumes of potential candidates
 - rate candidates on intelligence, warmth and commitment to work
 - recommend whether to hire or not, if yes, then what starting salary
- Same resumes, except treatment signalled motherhood, control didn't
- Less likely to hire mothers, with lower starting salary
- Statistical analysis revealed disadvantages for mothers largely explained by lower ratings in competence and commitment

Lab experiments

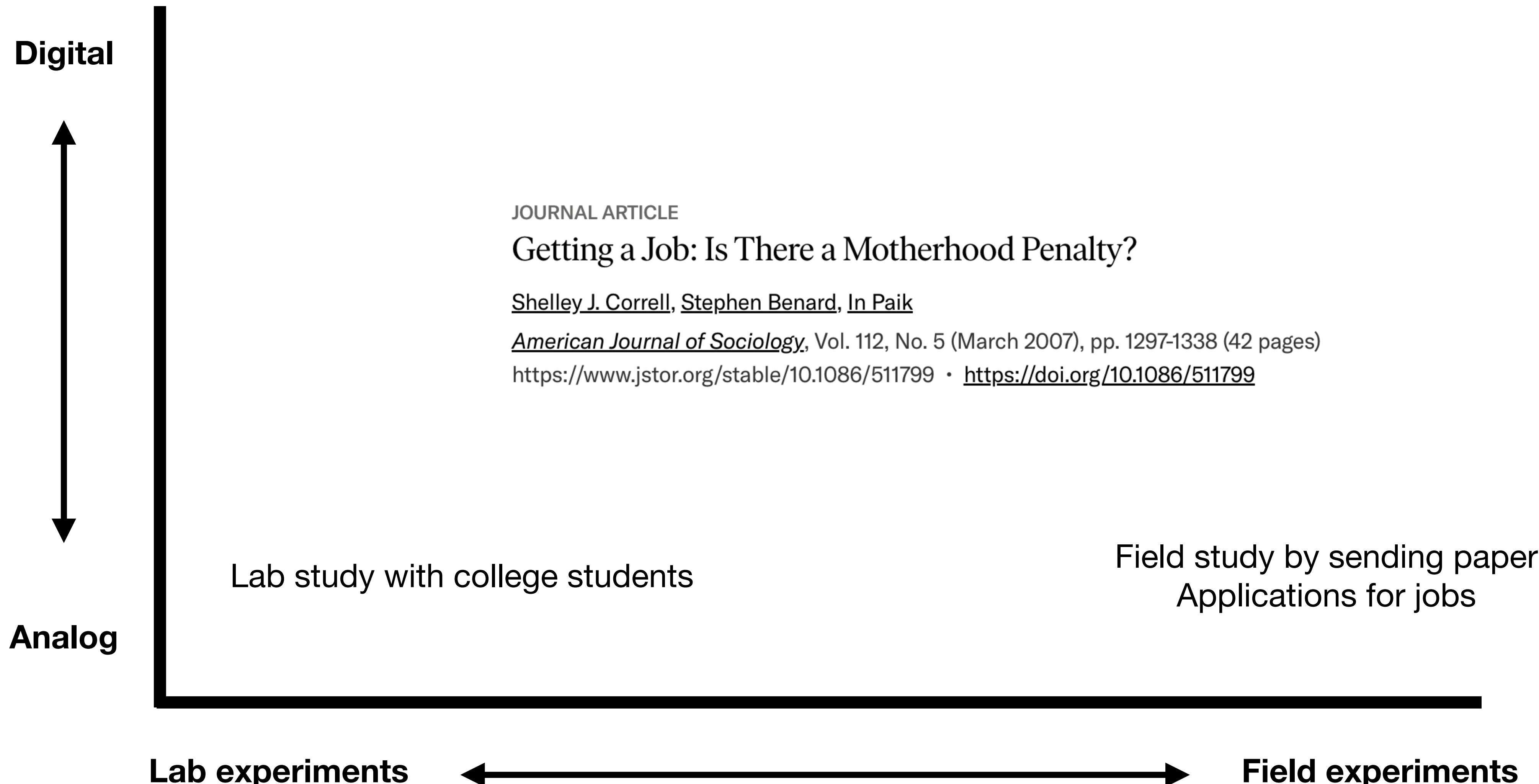


Field experiments

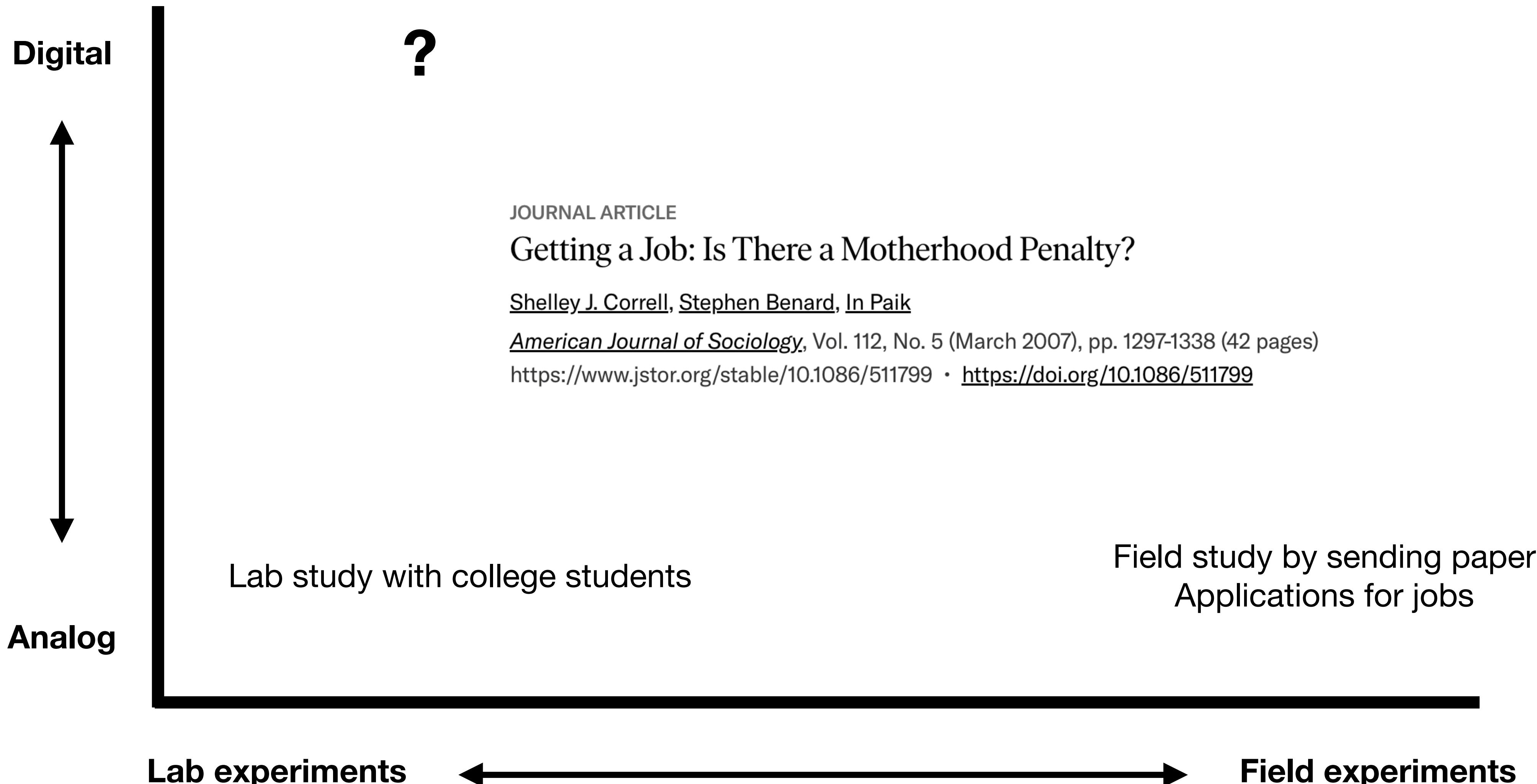
Dimensions of experiments



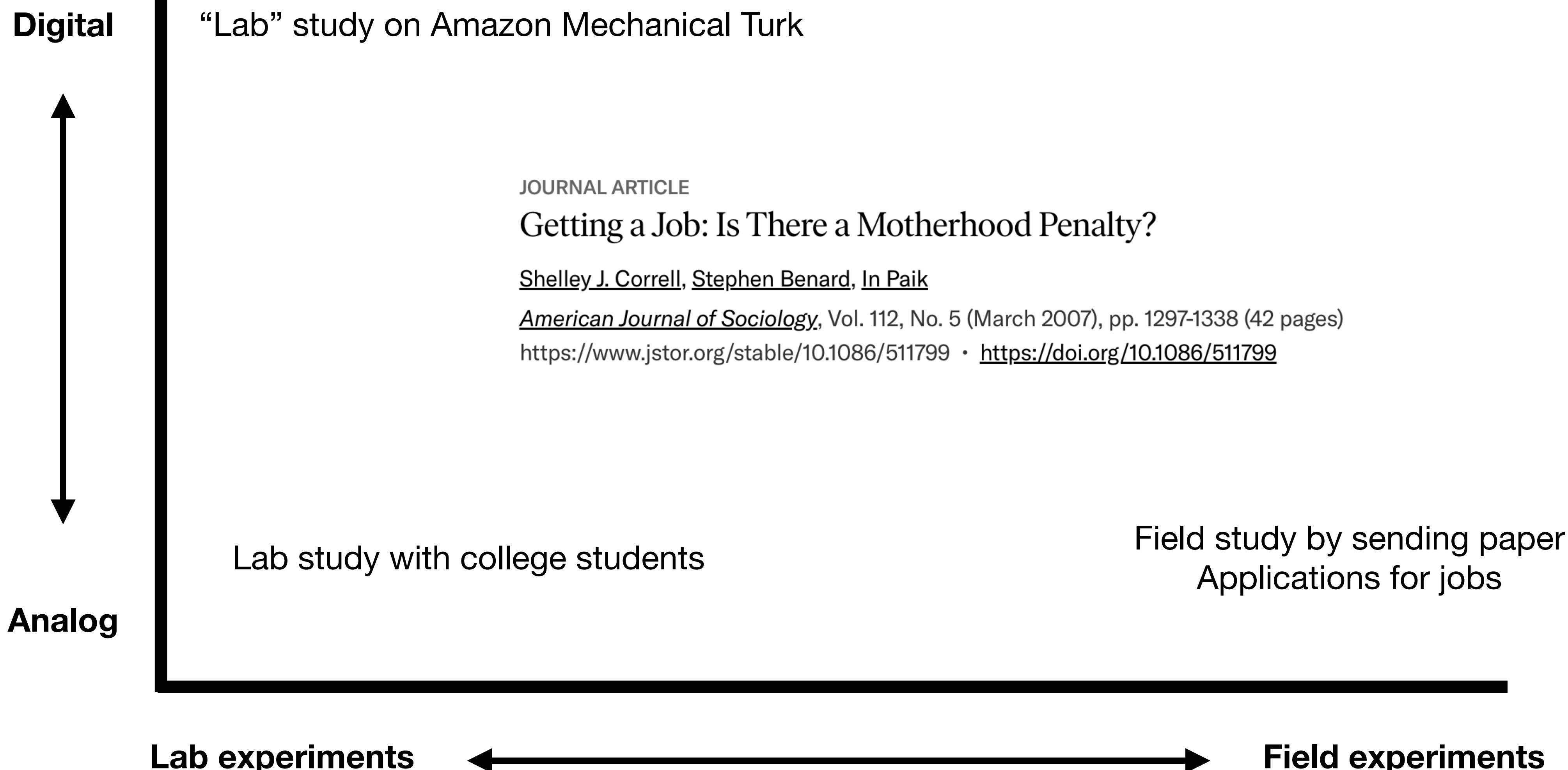
Dimensions of experiments



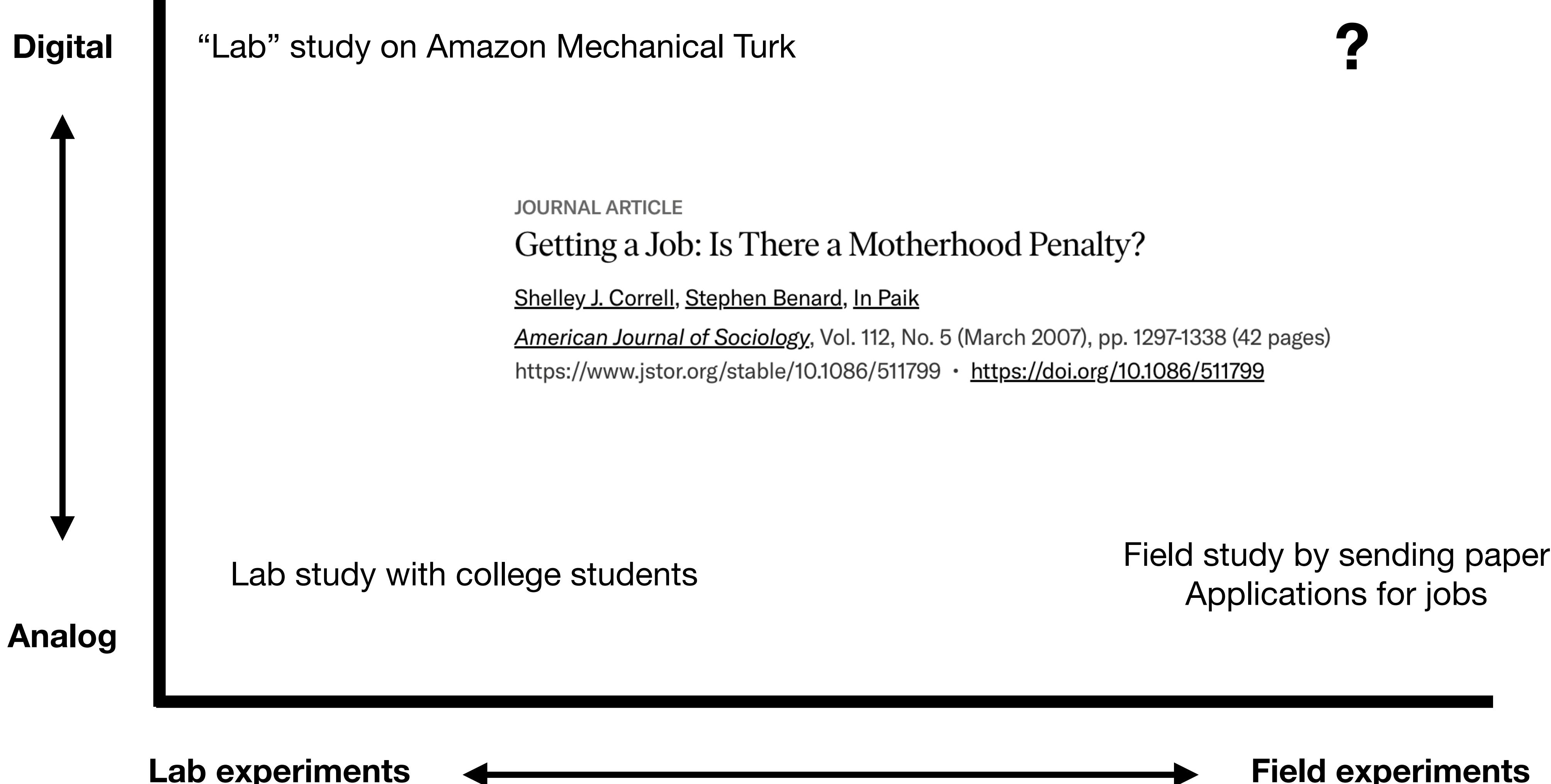
Dimensions of experiments



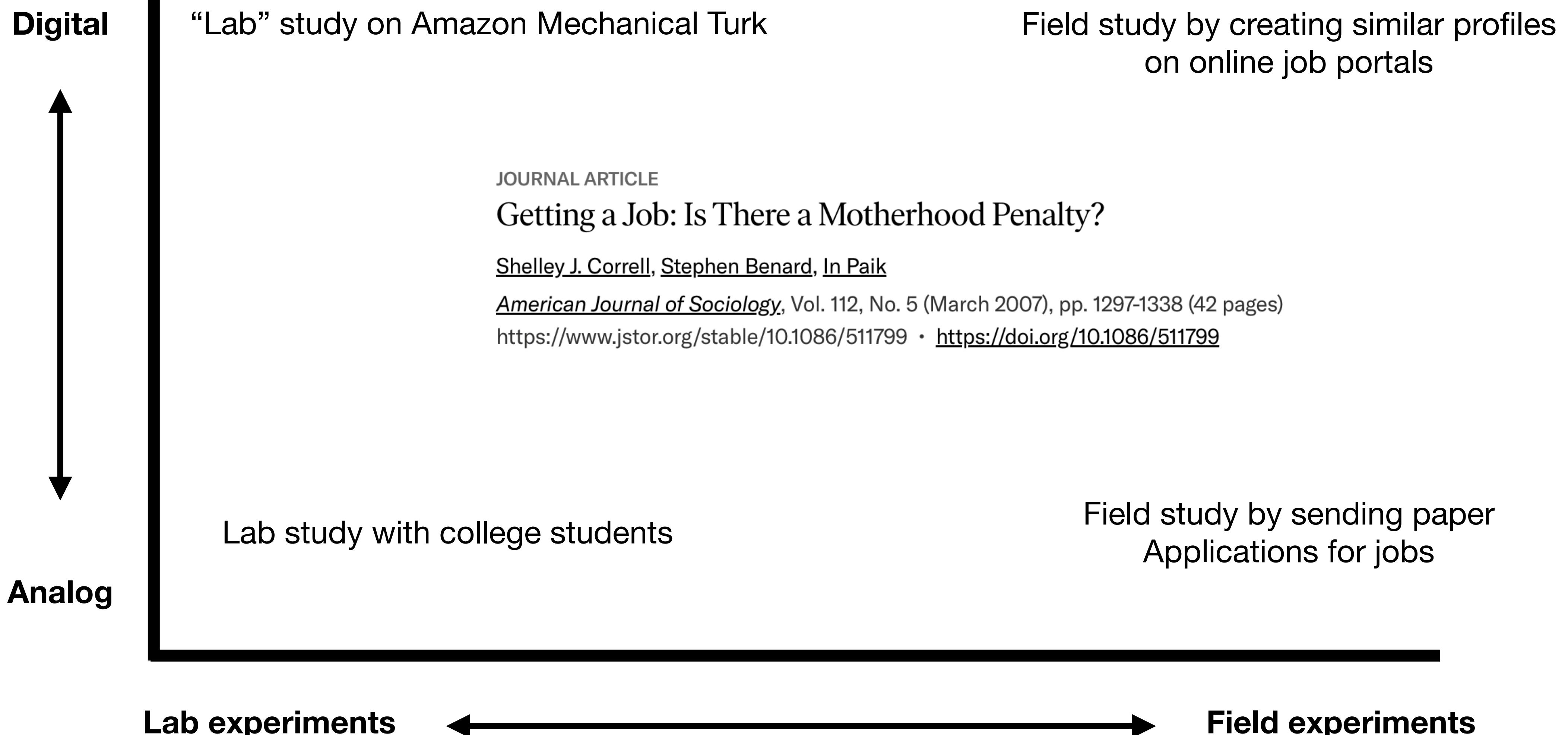
Dimensions of experiments



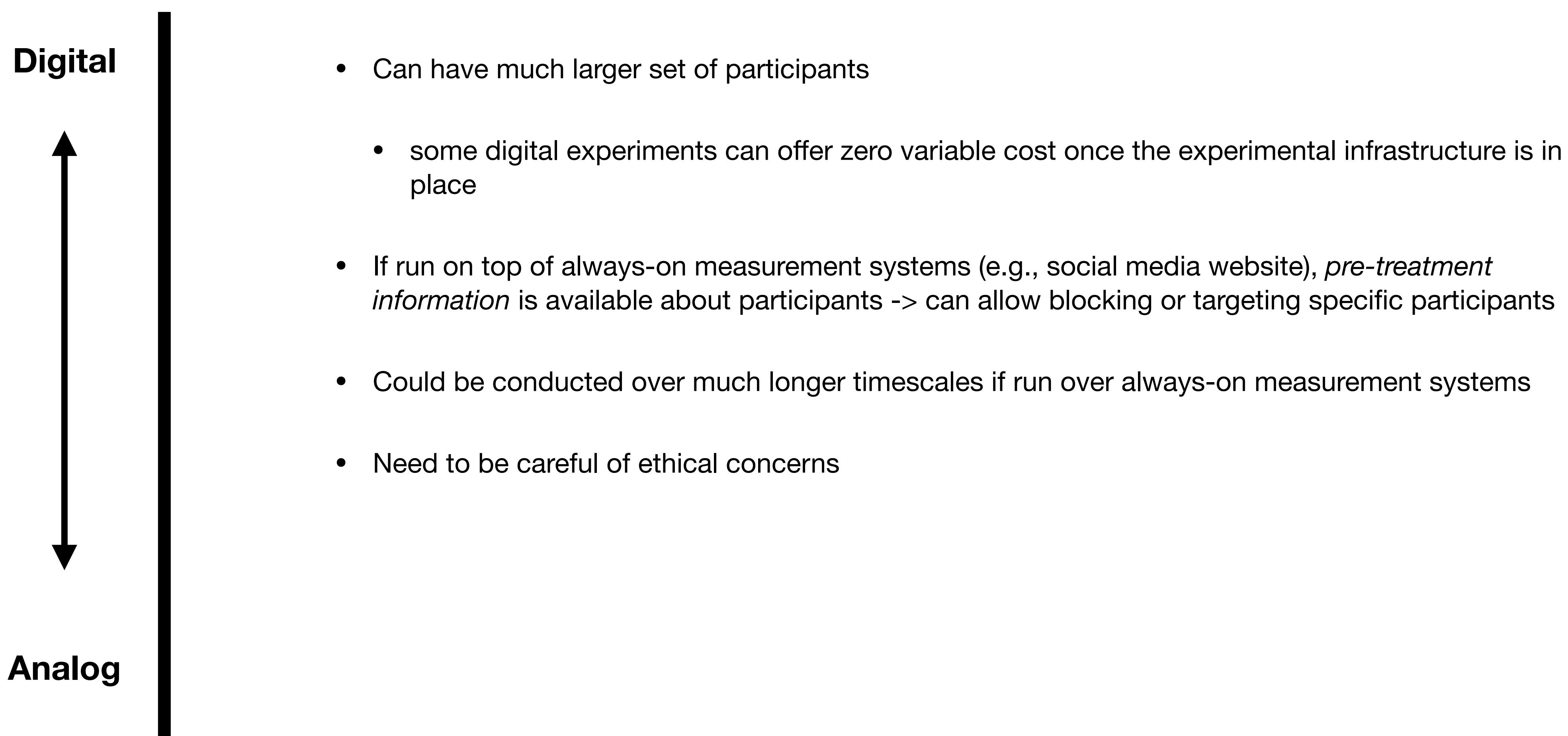
Dimensions of experiments



Dimensions of experiments



Dimensions of experiments



Optimisation experiments

vs.

Understanding experiments

Optimisation experiments
(e.g., A/B test)

vs.

Understanding experiments
(e.g., lab experiment to test an abstract theory of cognition)

Optimisation experiments

+

Understanding experiments

> *Psychol Sci*. 2007 May;18(5):429-34. doi: 10.1111/j.1467-9280.2007.01917.x.

The constructive, destructive, and reconstructive power of social norms

P Wesley Schultz¹, Jessica M Nolan, Robert B Cialdini, Noah J Goldstein, Vladas Griskevicius

Affiliations + expand

PMID: 17576283 DOI: [10.1111/j.1467-9280.2007.01917.x](https://doi.org/10.1111/j.1467-9280.2007.01917.x)

<https://pubmed.ncbi.nlm.nih.gov/17576283/>

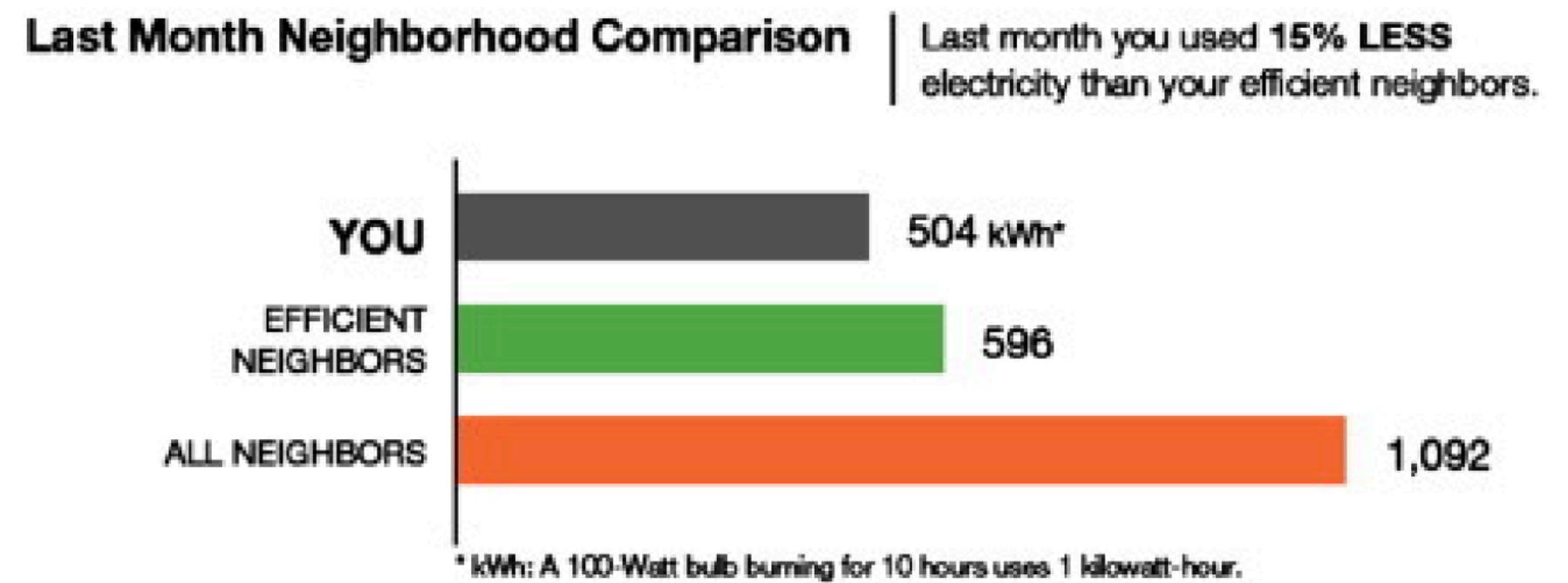
The constructive, destructive, and reconstructive power of social norms

P Wesley Schultz¹, Jessica M Nolan, Robert B Cialdini, Noah J Goldstein, Vladas Griskevicius

Affiliations + expand

PMID: 17576283 DOI: 10.1111/j.1467-9280.2007.01917.x

Decrease electricity usage through social norms by running an analog field experiment



Figures from Allcott (2011)

Convince people to reduce electricity consumption by showing them what their neighbours do

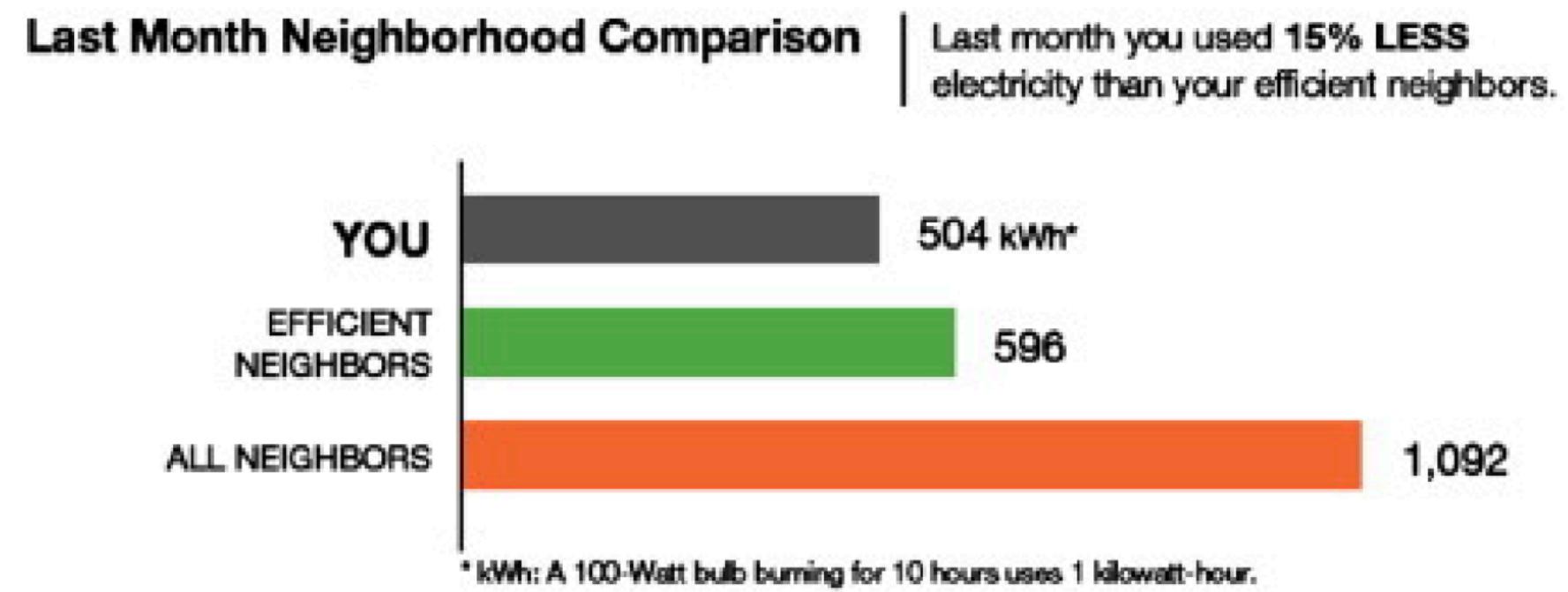
The constructive, destructive, and reconstructive power of social norms

P Wesley Schultz¹, Jessica M Nolan, Robert B Cialdini, Noah J Goldstein, Vladas Griskevicius

Affiliations + expand

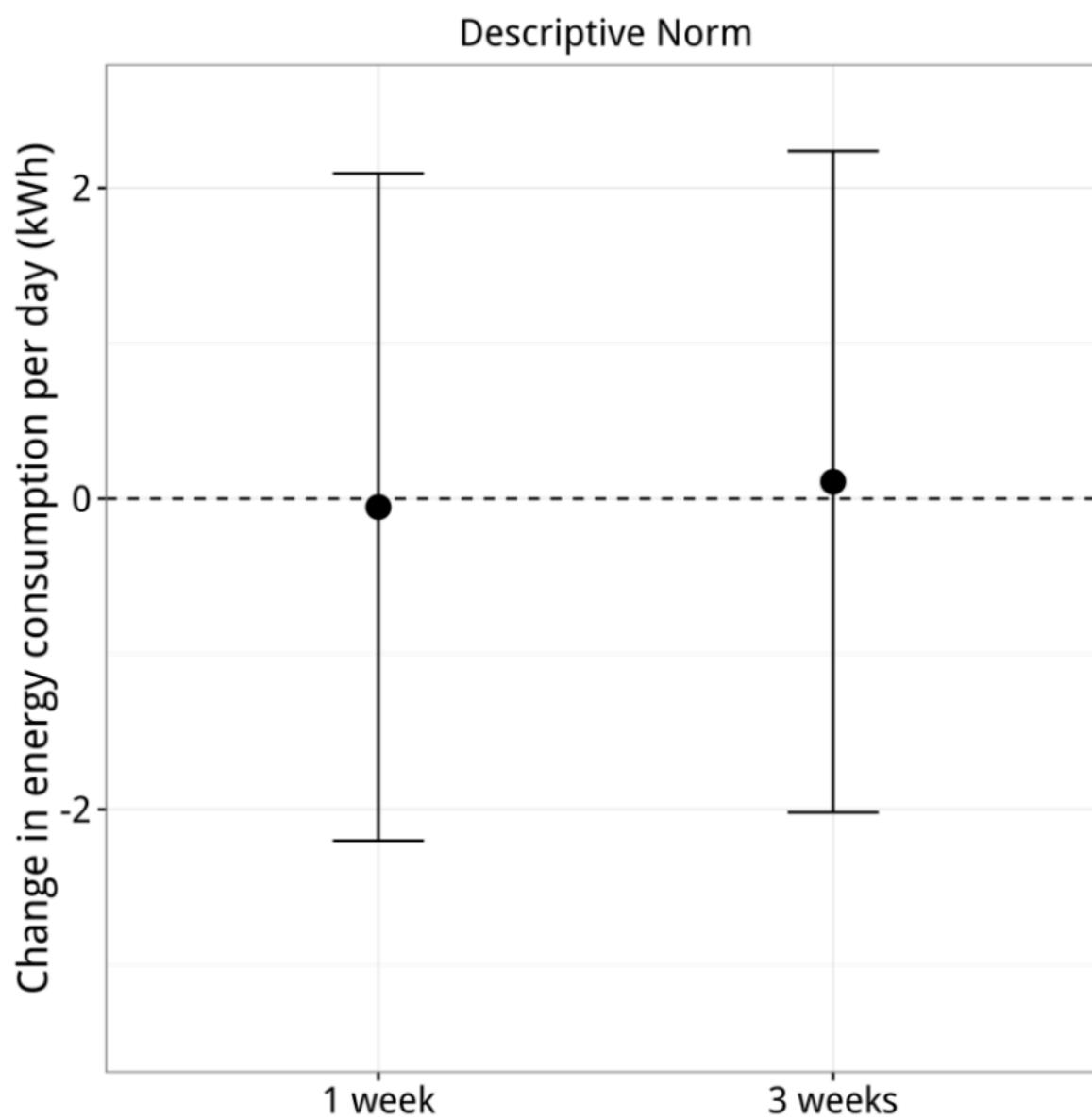
PMID: 17576283 DOI: 10.1111/j.1467-9280.2007.01917.x

Decrease electricity usage through social norms by running an analog field experiment



Figures from Allcott (2011)

Convince people to reduce electricity consumption by showing them what their neighbours do



- Zero average change in electricity consumption
- Seems treatment didn't work

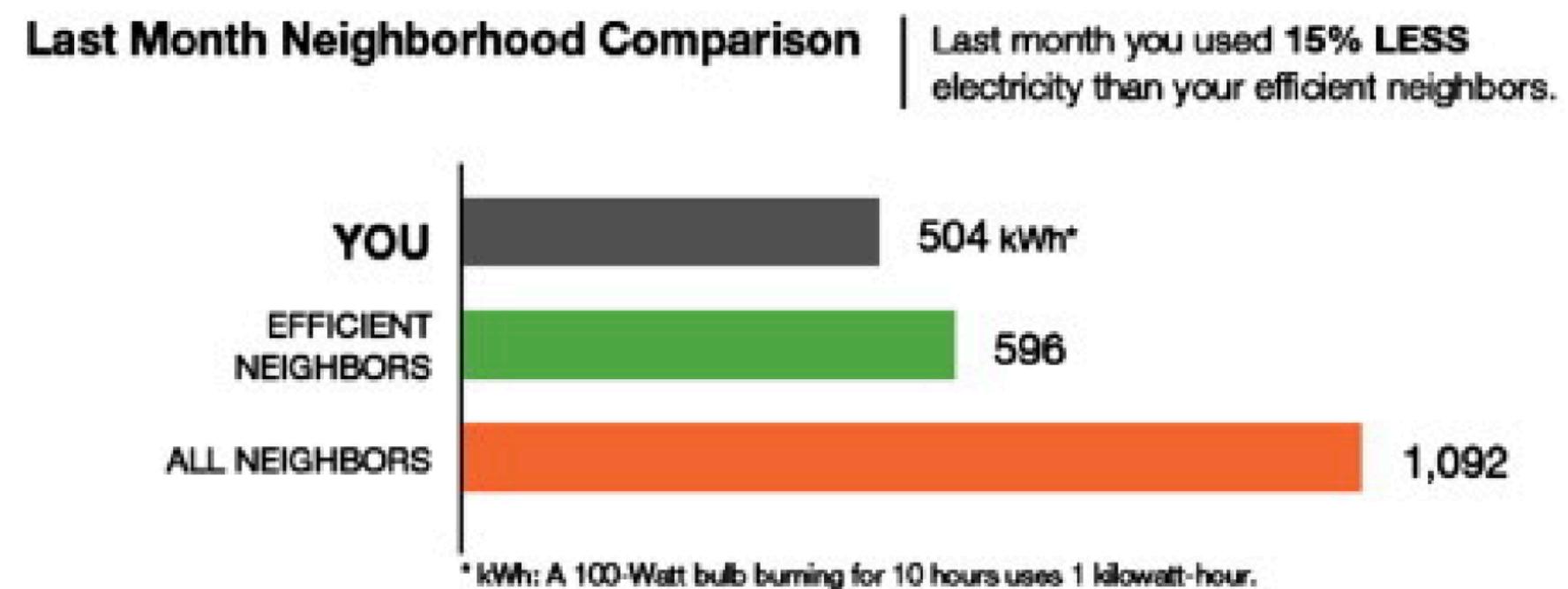
The constructive, destructive, and reconstructive power of social norms

P Wesley Schultz¹, Jessica M Nolan, Robert B Cialdini, Noah J Goldstein, Vladas Griskevicius

Affiliations + expand

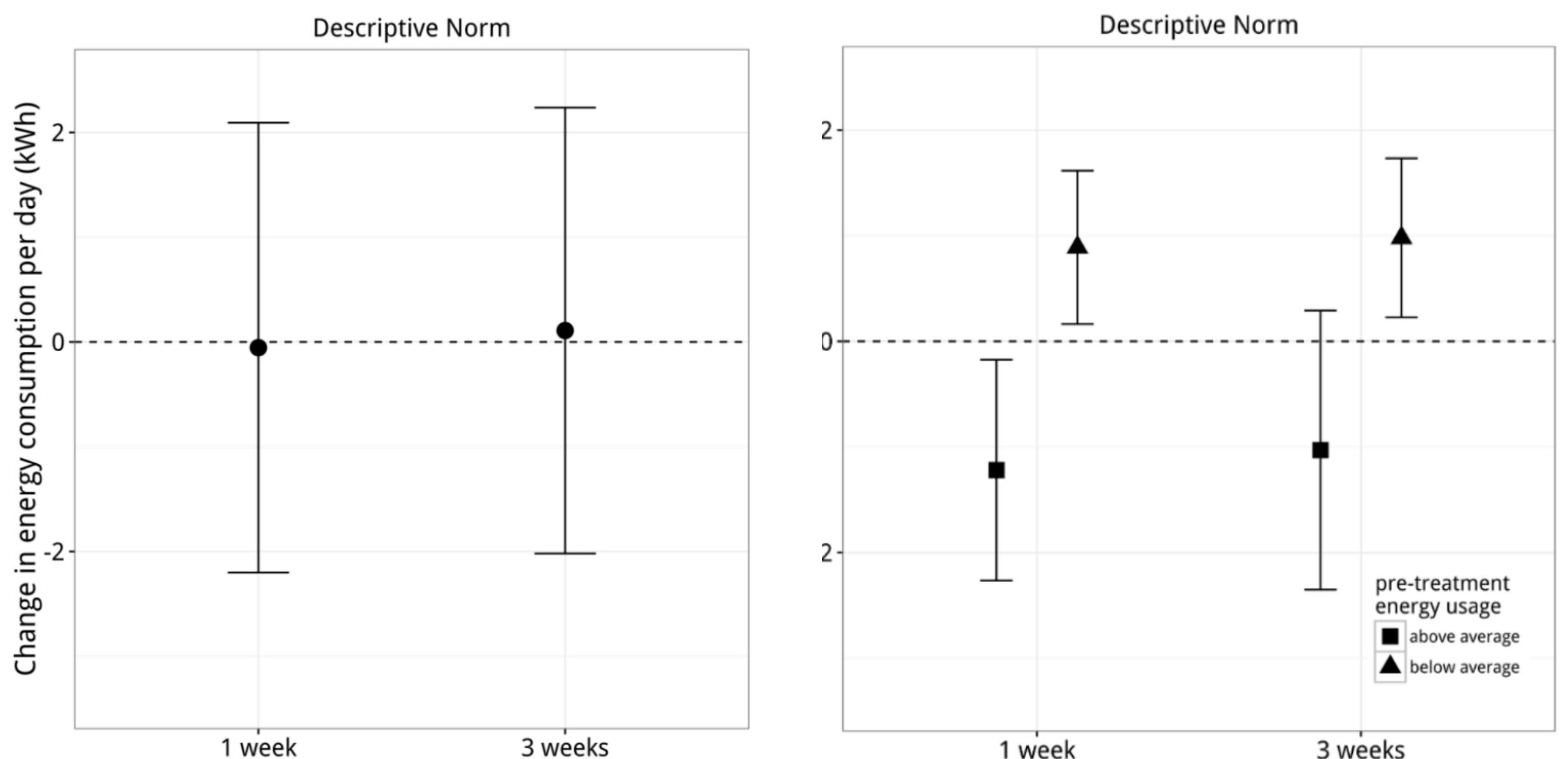
PMID: 17576283 DOI: 10.1111/j.1467-9280.2007.01917.x

Decrease electricity usage through social norms by running an analog field experiment



Figures from Allcott (2011)

Convince people to reduce electricity consumption by showing them what their neighbours do



- Zero average change in electricity consumption
- Seems treatment didn't work
- Hypothesised apriori that treatment would have 2 offsetting effects -
 - Used too much electricity, decreased it
 - Used less electricity, increased it (boomerang effect)

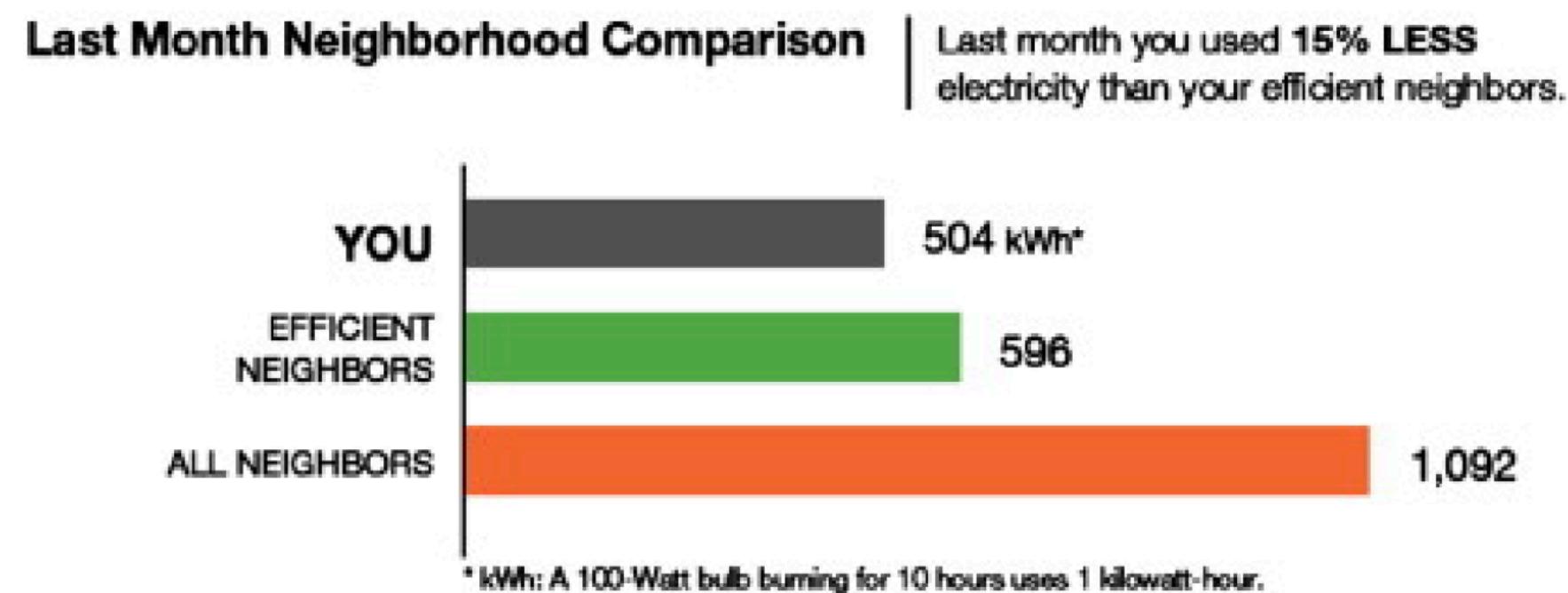
The constructive, destructive, and reconstructive power of social norms

P Wesley Schultz¹, Jessica M Nolan, Robert B Cialdini, Noah J Goldstein, Vladas Griskevicius

Affiliations + expand

PMID: 17576283 DOI: 10.1111/j.1467-9280.2007.01917.x

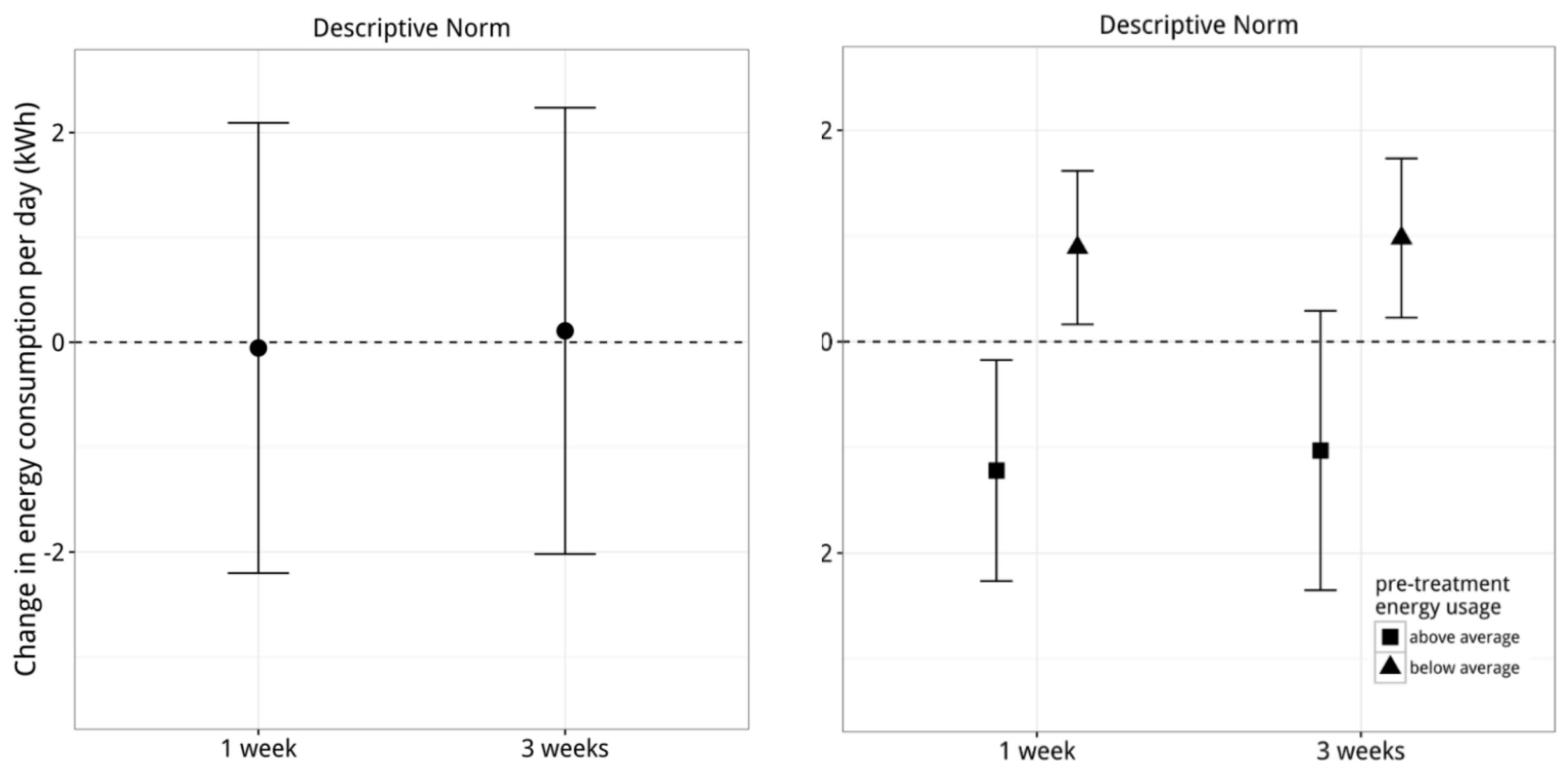
Decrease electricity usage through social norms by running an analog field experiment



Figures from Allcott (2011)

Convince people to reduce electricity consumption by showing them what their neighbours do

Heterogeneity -> what looks like non effect is multiple effects being averaged together



- Zero average change in electricity consumption
- Seems treatment didn't work
- Hypothesised apriori that treatment would have 2 offsetting effects -
 - Used too much electricity, decreased it
 - Used less electricity, increased it (boomerang effect)

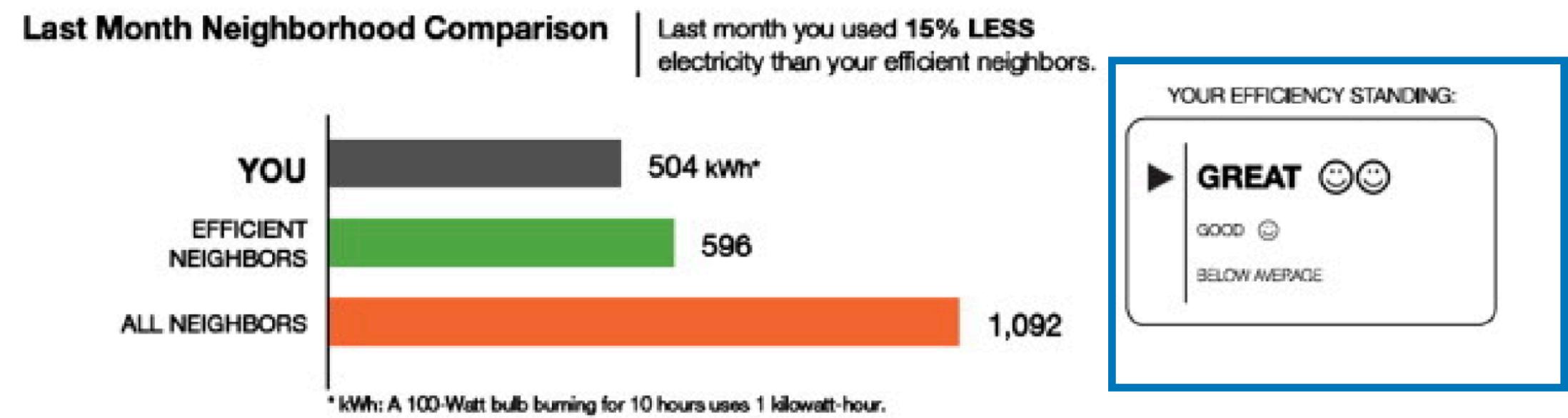
The constructive, destructive, and reconstructive power of social norms

P Wesley Schultz¹, Jessica M Nolan, Robert B Cialdini, Noah J Goldstein, Vladas Griskevicius

Affiliations + expand

PMID: 17576283 DOI: 10.1111/j.1467-9280.2007.01917.x

Decrease electricity usage through social norms by running an analog field experiment



Figures from Allcott (2011)

Not just descriptive norm, but also injunctive norm of what people consider more socially desirable

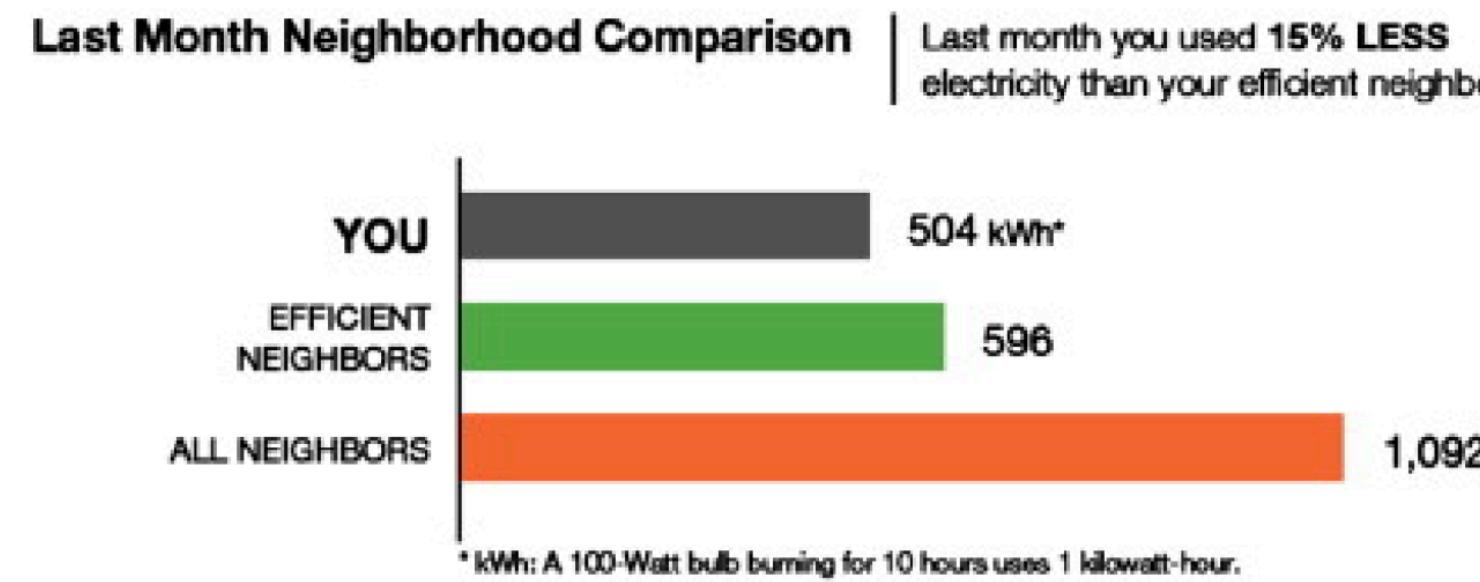
The constructive, destructive, and reconstructive power of social norms

Decrease electricity usage through social norms by running an analog field experiment

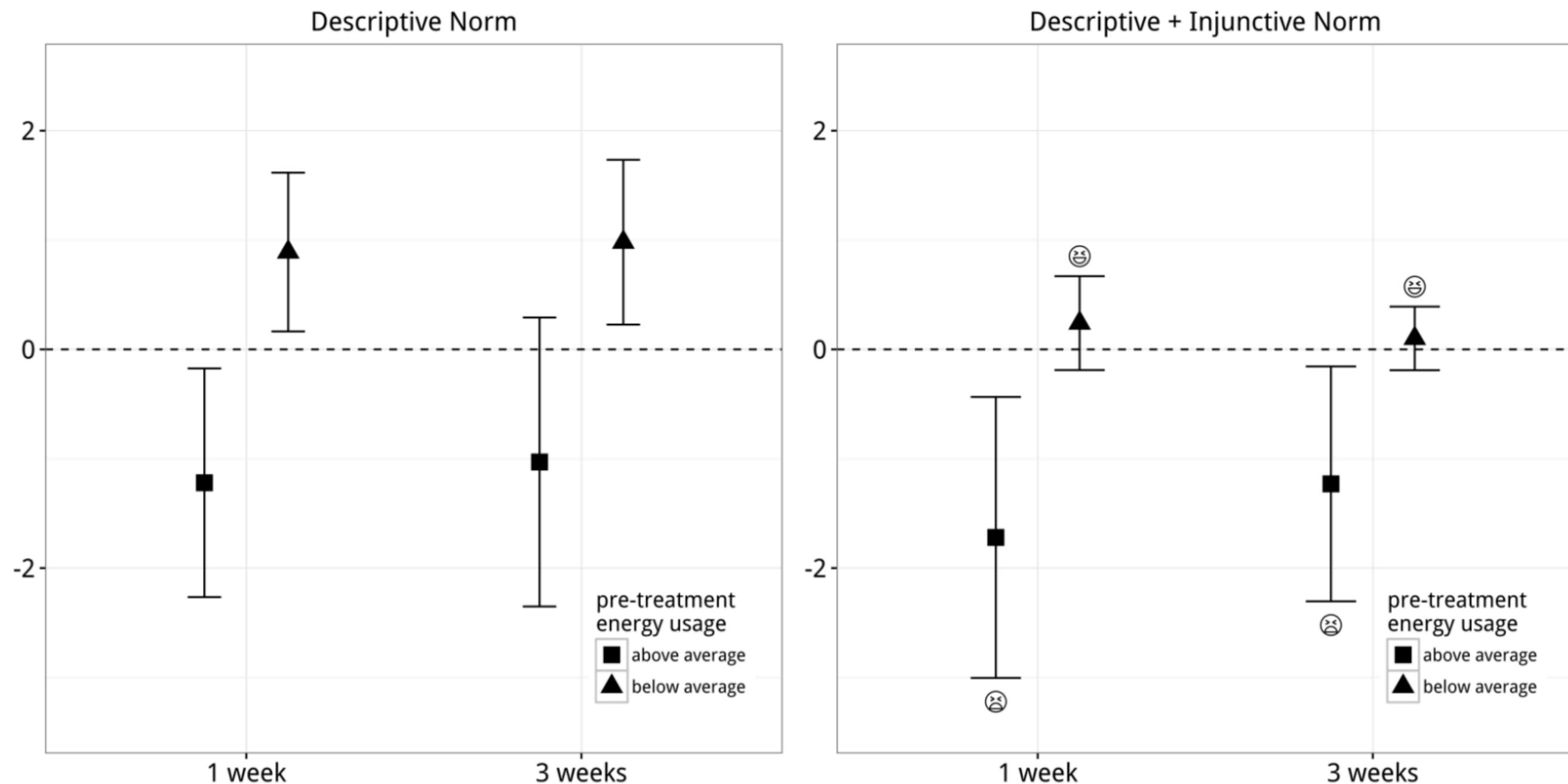
P Wesley Schultz¹, Jessica M Nolan, Robert B Cialdini, Noah J Goldstein, Vladas Griskevicius

Affiliations + expand

PMID: 17576283 DOI: 10.1111/j.1467-9280.2007.01917.x



Figures from Allcott (2011)



Not just descriptive norm, but also injunctive norm of what people consider more socially desirable

- Reduced the boomerang effect
- Theory of norms allowed them to design a treatment that would be more effective

Concepts for designing richer experiments

- Validity
- Heterogeneity of treatment effects
- Mechanisms

Validity

- How much do the results of an experiment support a more general conclusion?
- Checklist of the main types of ways how your experiments can go wrong
- Four main types of validity:
 - Statistical conclusion validity
 - Internal validity
 - Construct validity
 - External validity

Statistical conclusion validity

- Whether the statistical analysis of the experiment was done correctly
- Similar for analog and digital experiments

Internal validity

- Whether the experimental procedures were performed correctly
- Did your experiment do what you think it did
 - Were you able to deliver the treatment correctly
 - Was there any information leak between the treatment and control
- Reduced concerns in digital age -
 - Easier to ensure treatment is delivered to treatment group
 - Easier to measure outcomes for all participants

Construct validity

- Whether the data collected matches the theoretical construct of interest
- Construct are abstract concepts of interest (e.g., bias, sexism) that we want to reason about
- Constructs usually don't have clear definitions or measures therefore are very hard to operationalise
- Have you operationalised your construct well?
- Big concern in digital experiments

External validity

- Whether the results of the experiment can be generalised to other situations
- Would the results hold for other groups of participants or different ways of operationalising the construct
- Analog experiments - usually reasoned about theoretically
- Digital experiments - assess external validity empirically, since experiments are cheaper to run

Heterogeneous treatment effects

- Experiments measure average effect, but the same treatment can have different effect on different kinds of people
 - Electricity usage experiment - consumption went up for one group and down for another
- Large heterogeneous set of participants
 - Lots of participants (100 vs. 100,000)
 - Often lots of pre-treatment information about participants, use it to look for heterogeneity
 - Can help understand how a treatment works, how it can be improved, and how it can be targeted to those most likely to benefit

Heterogeneous treatment effects

- Fishing - Split data into lots of groups and look where effects are bigger/smaller, and if you do enough splits you will find some groups for which effects are bigger/smaller
- Avoid fishing
 - Preregistration - say ahead of time which groups you are interested in, prior to treatment
 - Split the data into two, in first half look for heterogeneity in treatment effects, then preregister the heterogeneity of interest and then measure it in the second half not yet looked at
 - New methods available for automatically looking at heterogeneous treatment effects while avoiding the problems of fishing

Mechanisms

- Example: Scurvy (sailors suffered greatly from it), found that citrus prevents scurvy
- Experiments to test whether citrus prevents scurvy, but could not answer why
- More research to find that the mechanism through which citrus prevents scurvy is vitamin C
- Mechanism important to find more efficient solutions (e.g., other longer lasting fruits with vitamin C, vitamin C pills)
- Not always easy to find out the mechanism,
- In digital age it is easier to specifically design experiments to test for mechanisms – Ludwig et al 2011, Imai et all 2012, Pirlott and MacKinnon 2016

Making digital experiments happen

Making digital experiments happen

Cost

Money to build experiment and pay participants
Time and effort to do so

Making digital experiments happen

Cost Control

Extent to which you can control the environment in
which people are doing your experiment

Making digital experiments happen

Cost Control Realism

To what extent your experiment involves naturally occurring treatments, environments, participants

Making digital experiments happen

Cost Control Realism Ethics

To what extent you are likely to run into ethical challenges with your approach

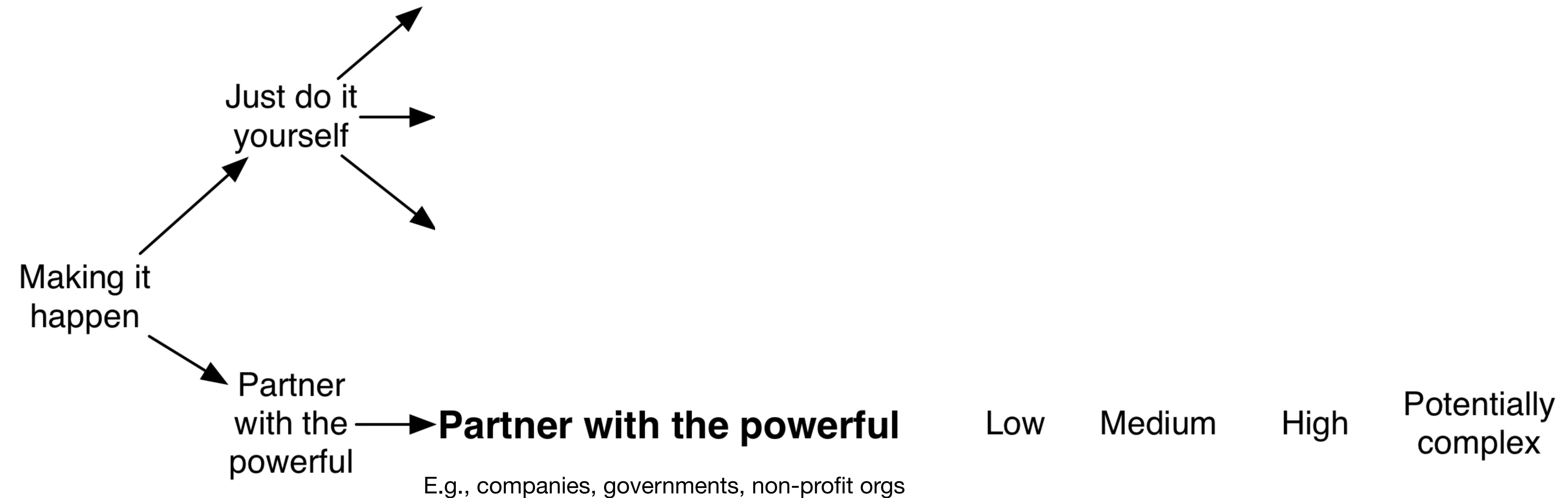
Making digital experiments happen

Cost Control Realism Ethics



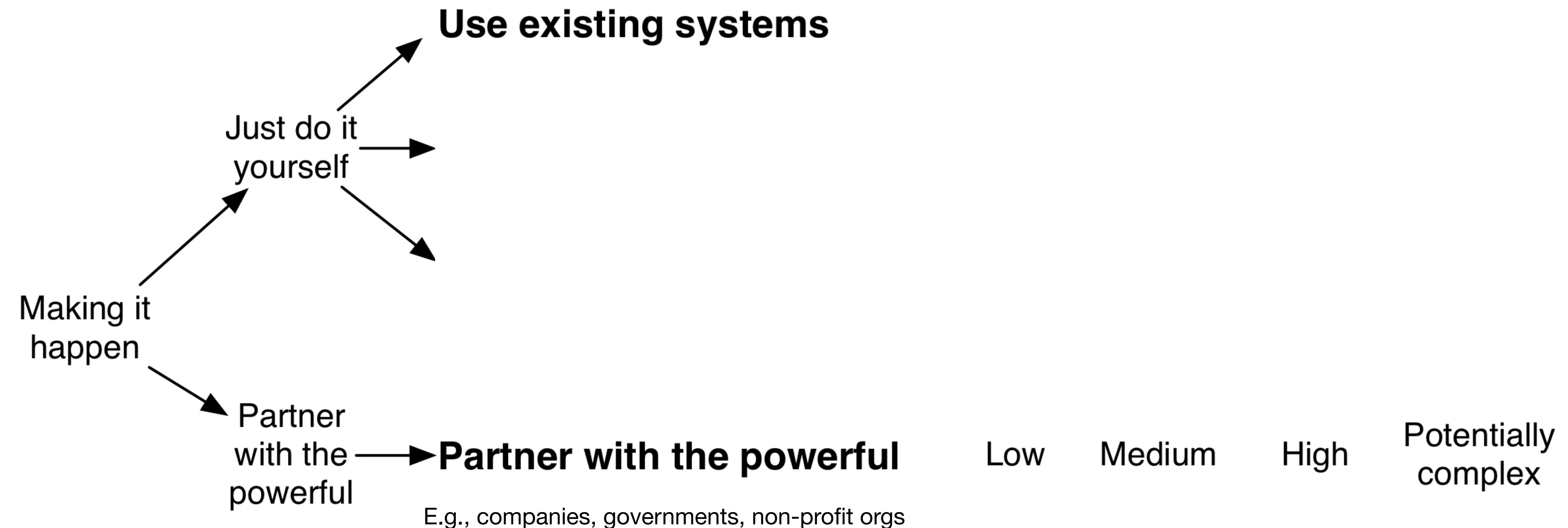
Making digital experiments happen

Cost Control Realism Ethics



Making digital experiments happen

Cost Control Realism Ethics



Use an existing system



Feature Article

The Visible Hand: Race and Online Market Outcomes

Jennifer L. Doleac ✉, Luke C.D. Stein

First published: 24 August 2013 | <https://doi.org/10.1111/ecoj.12082> | Citations: 123

Use an existing system



OPEN ACCESS Freely available online

Experimental Study of Informal Rewards in Peer Production

Michael Restivo*, Arnout van de Rijt

Department of Sociology, State University of New York at Stony Brook, New York, New York, United States of America

Abstract

We test the effects of informal rewards in online peer production. Using a randomized, experimental design, we assigned editing awards or “barnstars” to a subset of the 1% most productive Wikipedia contributors. Comparison with the control group shows that receiving a barnstar increases productivity by 60% and makes contributors six times more likely to receive additional barnstars from other community members, revealing that informal rewards significantly impact individual effort.

The
**ECONOMIC
JOURNAL**



Feature Article

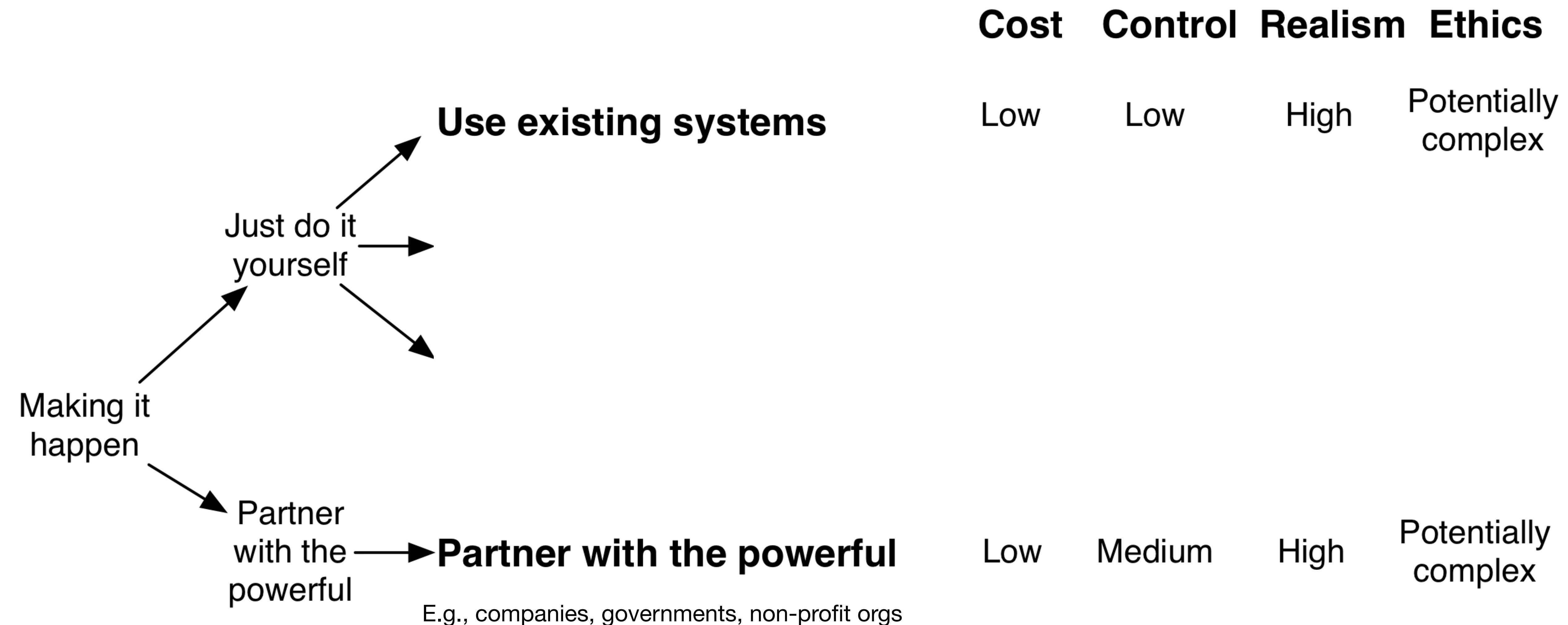
The Visible Hand: Race and Online Market Outcomes

Jennifer L. Doleac Luke C.D. Stein

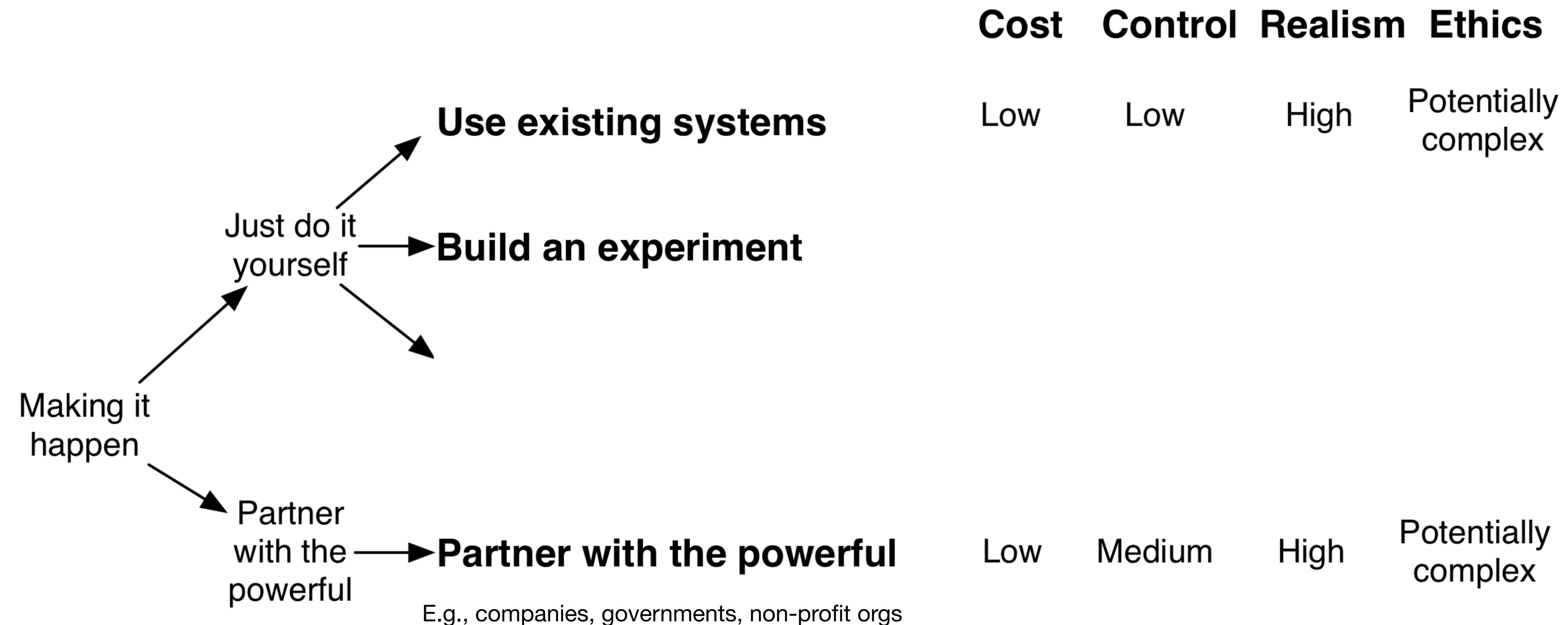
First published: 24 August 2013 | <https://doi.org/10.1111/ecoj.12082> | Citations: 123

PLoS one

Making digital experiments happen



Making digital experiments happen



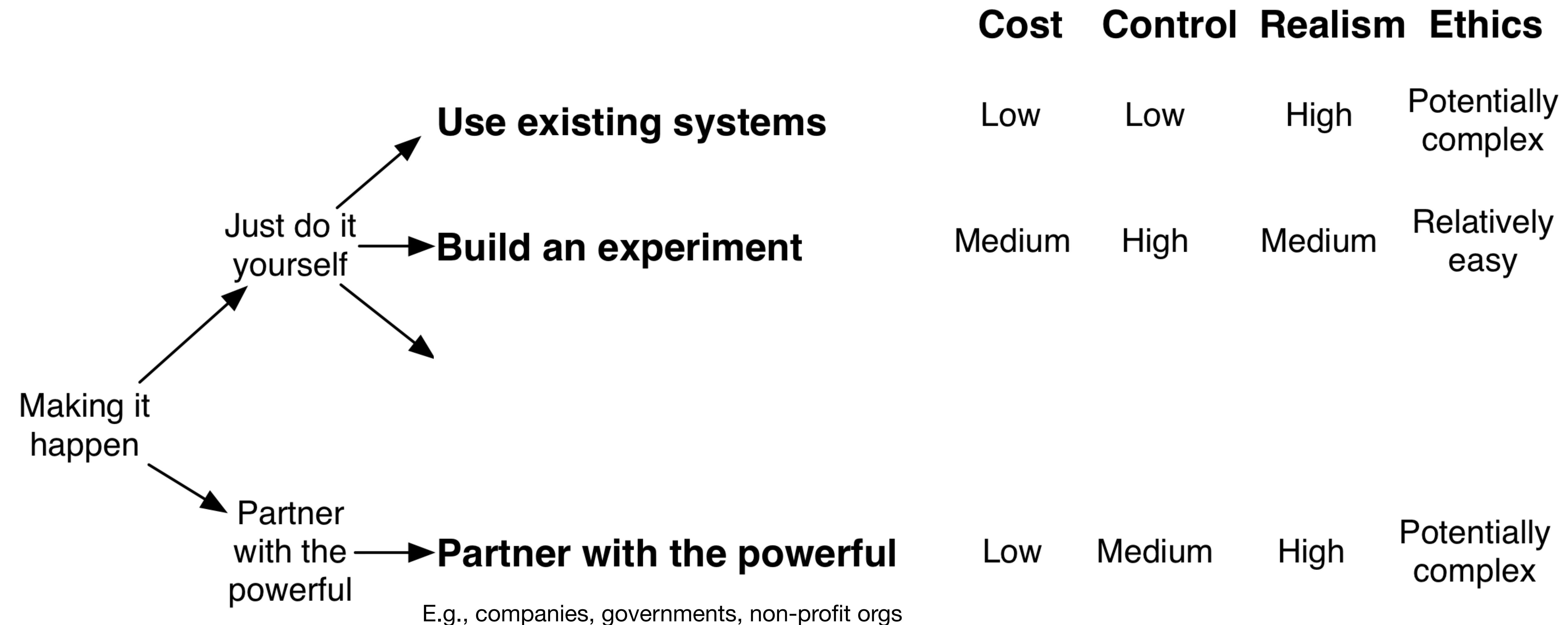
Build an experiment

- Built a digital field experiment to study effect of social network structure on spread of behavior
- Observe the same behavior spreading in populations that had different social network structures but were otherwise indistinguishable
- Online health community, 1500 participants recruited via ads on health websites
- Assigned “health buddies” on arrival - create different social network structures in different groups - random vs. Clustered networks
- Injected a new behavior into each network: the chance to register for a new website with additional health information. When one signed up, all of their health buddies received an email announcing this behavior.
- Found behavior spread further and faster in the clustered network than in the random network

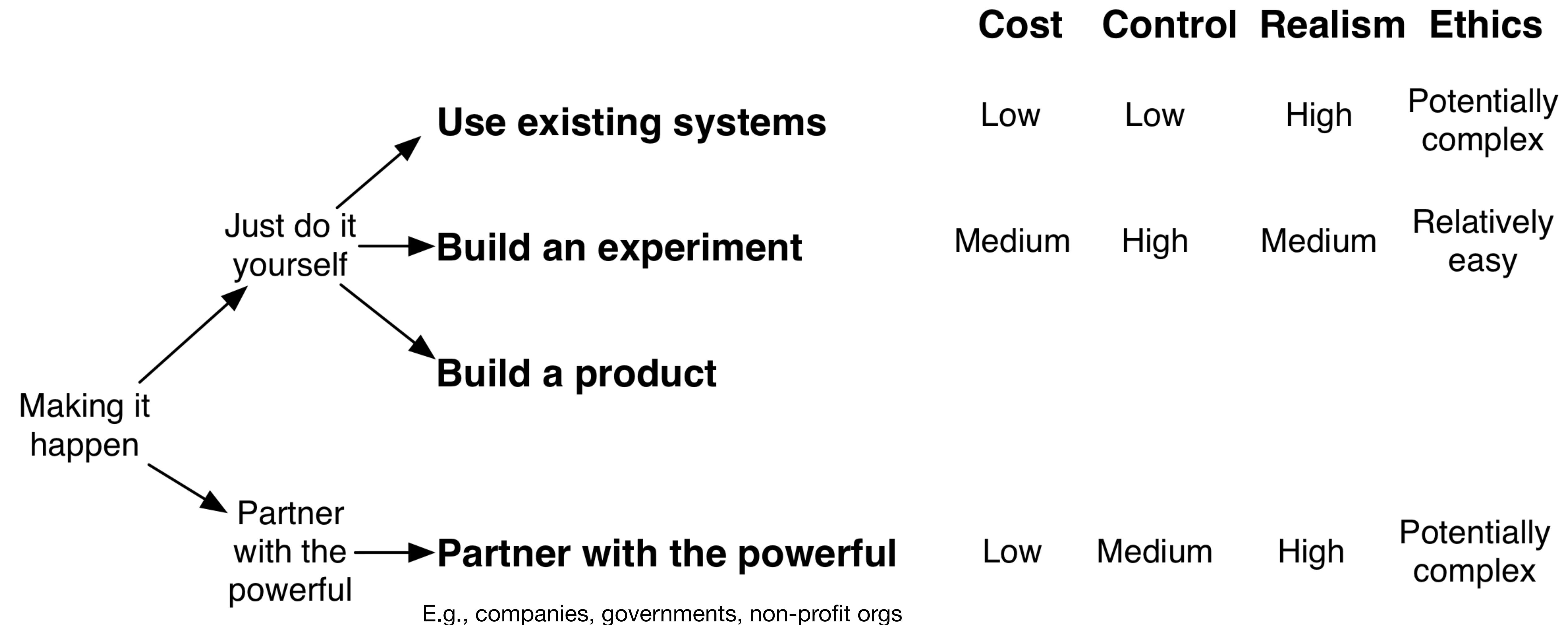
The Spread of Behavior in an Online Social Network Experiment

DAMON CENTOLA [Authors Info & Affiliations](#)

Making digital experiments happen



Making digital experiments happen



Build a product



recommendations

MovieLens helps you find movies you will like. Rate movies to build a custom taste profile, then MovieLens recommends other movies for you to watch.

A screenshot of the MovieLens website showing recommendation sections. The first section, "top picks", displays a grid of movie cards with titles like "Band of Brothers", "Casablanca", and "One Flew Over the Cuckoo's Nest". Below this is a row of five star rating icons. The second section, "recent releases", shows a grid of movie cards for films released in the last 90 days, including "Cantinflas", "Felony", and "What If".

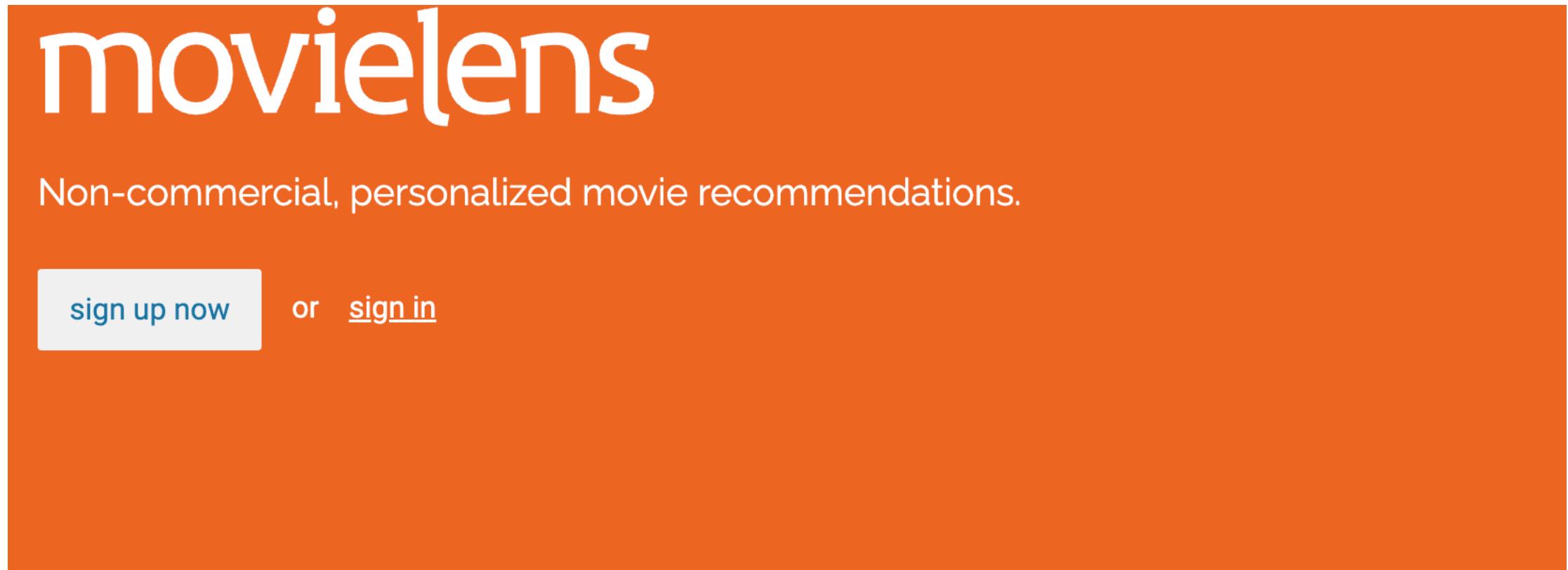
top picks see more
based on your ratings, MovieLens recommends these movies

Band of Brothers 2001 [R] 705 min	Casablanca 1942 [PG] 102 min	One Flew Over the Cuckoo's Nest 1975 [R] 133 min	The Lives of Others 2006 [R] 137 min	Sunset Boulevard 1950 [NR] 110 min	The Third Man 1949 [NR] 104 min	Path
--------------------------------------	---------------------------------	---	---	---------------------------------------	------------------------------------	------

recent releases see more
movies released in last 90 days that you haven't rated

Cantinflas 2014 [PG] 106 min	Felony 2014	What If 2014 [PG-13] 102 min	Frank 2014 [R] 96 min	Sin City: A Dame to Kill For 2014 [R] 102 min	If I Stay 2014 [PG-13] 106 min	Are We There Yet?
---------------------------------	----------------	---------------------------------	--------------------------	--	-----------------------------------	-------------------

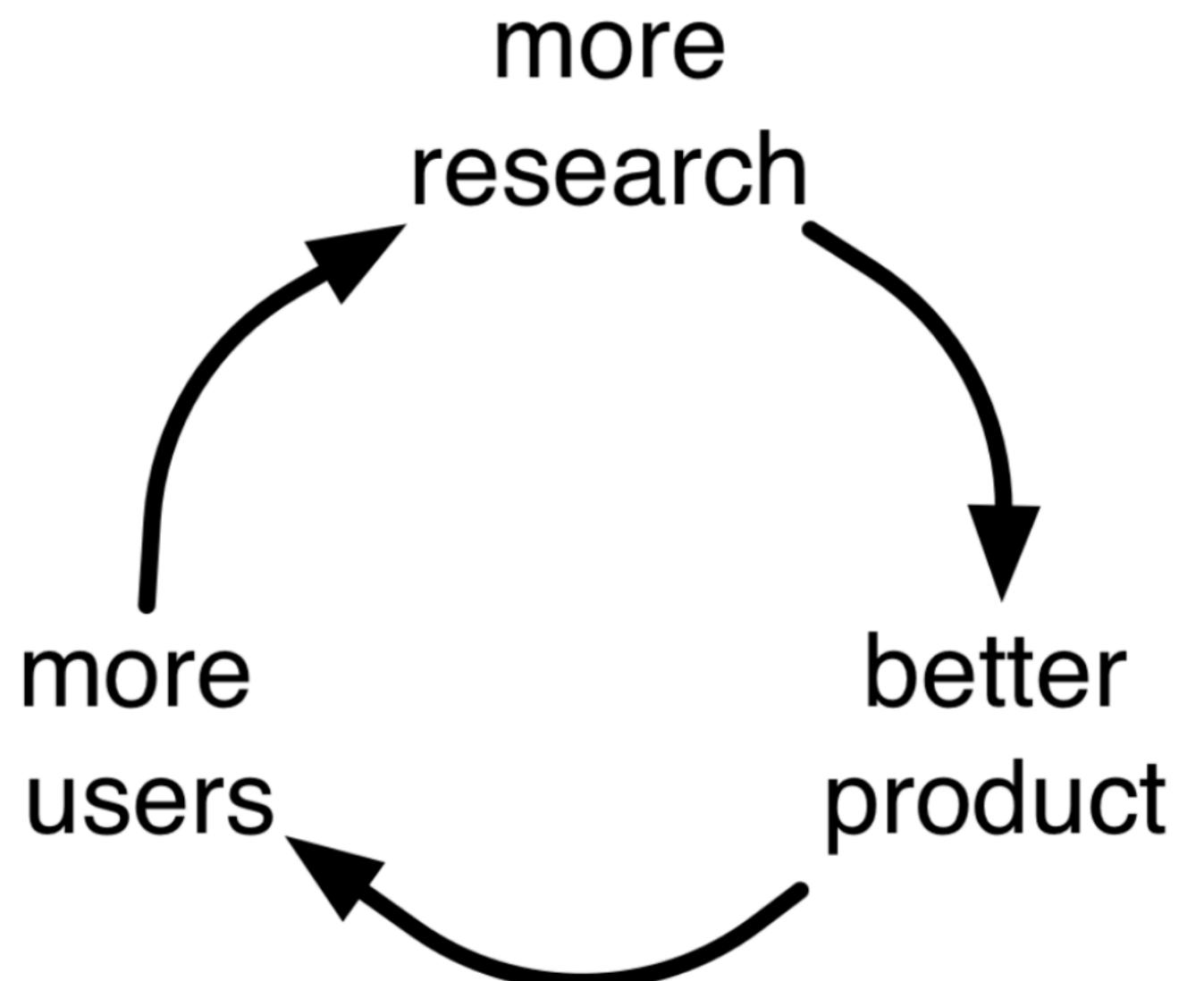
Build a product



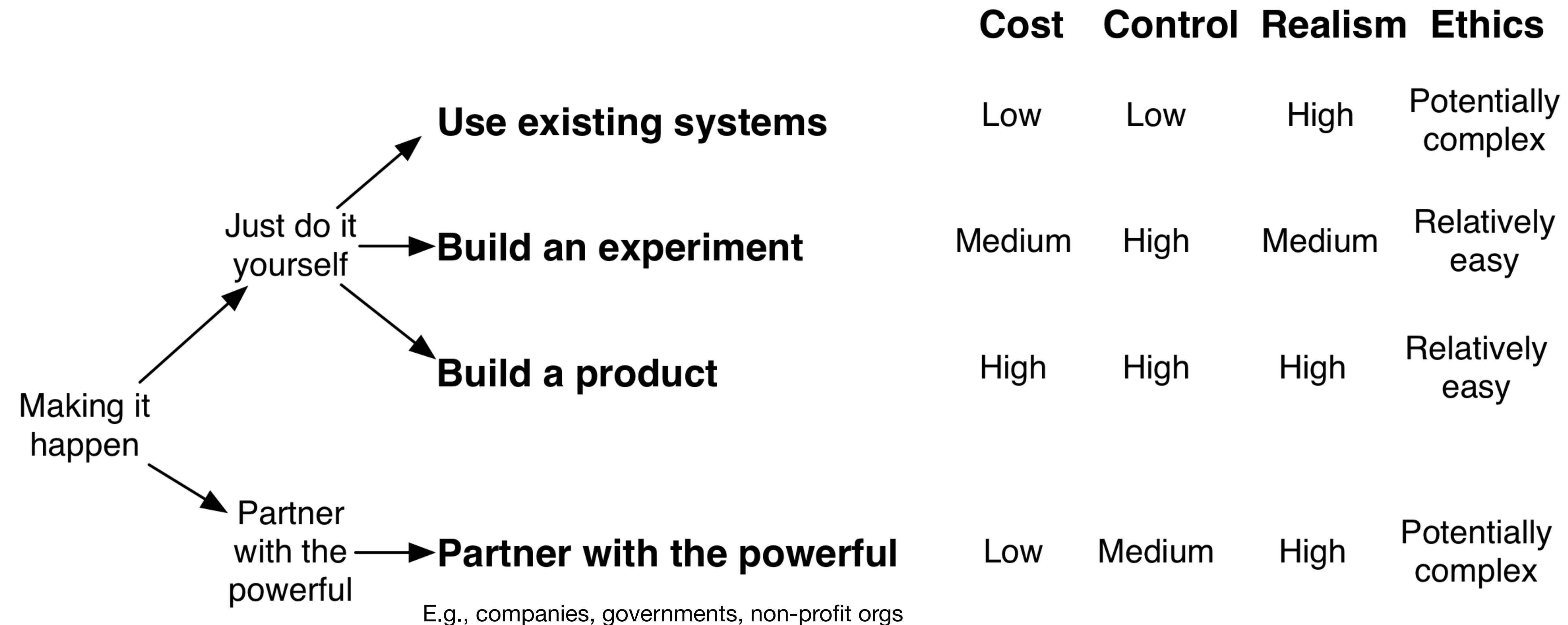
recommendations

MovieLens helps you find movies you will like. Rate movies to build a custom taste profile, then MovieLens recommends other movies for you to watch.

A screenshot of the MovieLens website showing recommendation sections. The top section is titled "top picks" and shows a grid of movie thumbnails with titles like "Band of Brothers", "Casablanca", and "One Flew Over the Cuckoo's Nest". Below it is a section titled "recent releases" which shows movies released in the last 90 days.

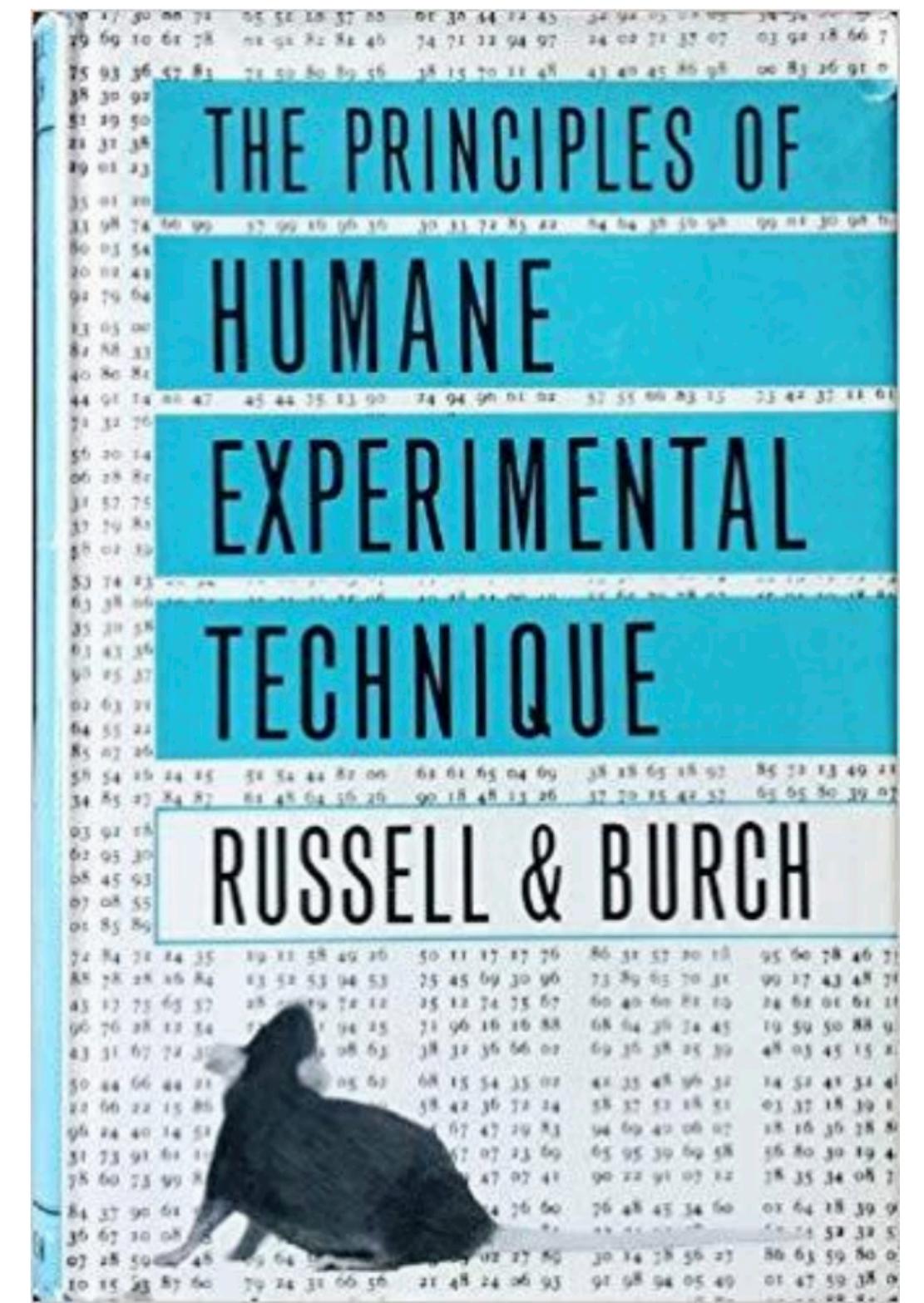


Making digital experiments happen



Building ethics in experiment design

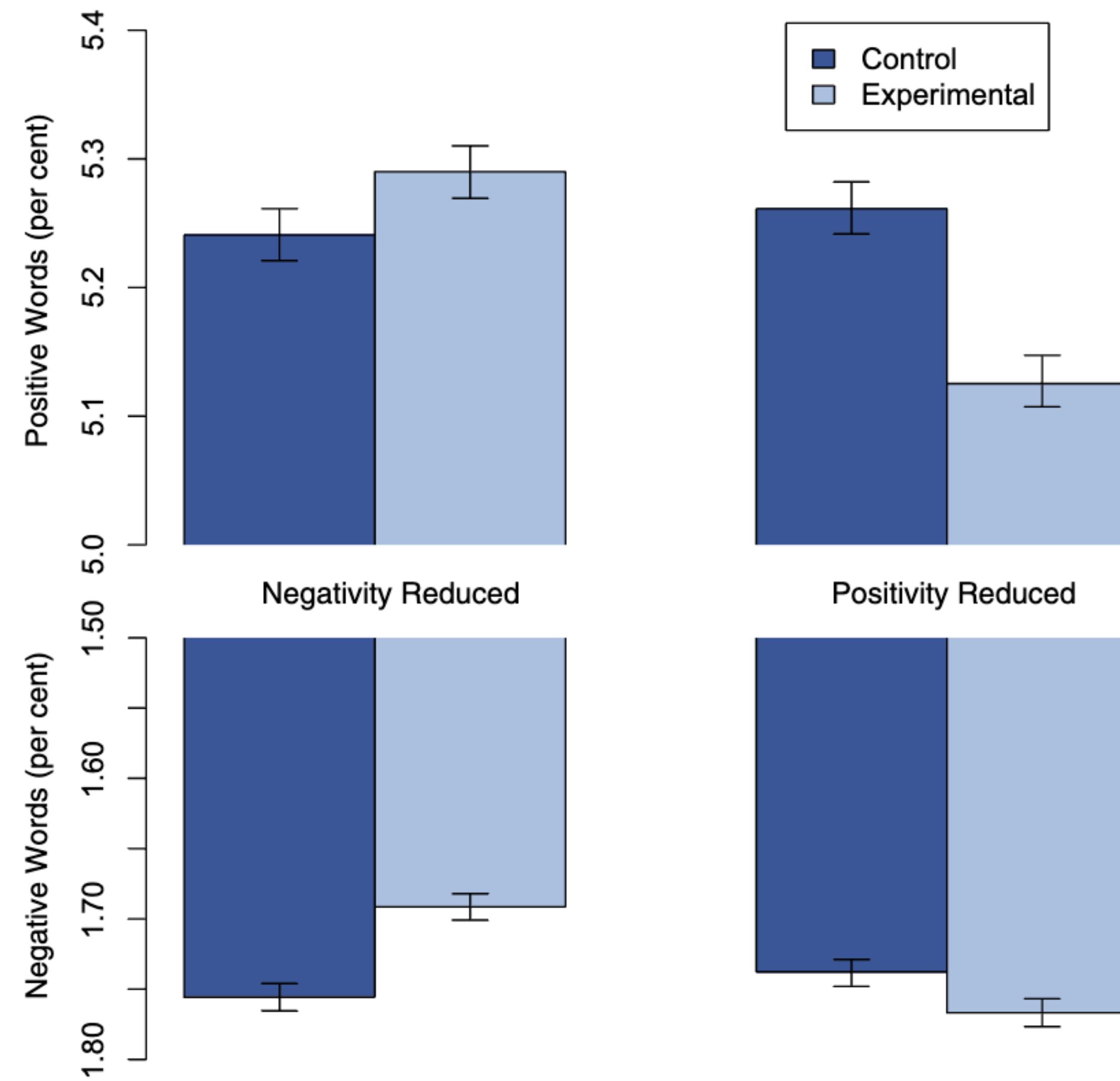
- The 3 R's
 - Replace
 - Refine
 - Reduce



Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer^{a,1}, Jamie E. Guillory^{b,2}, and Jeffrey T. Hancock^{b,c}

^aCore Data Science Team, Facebook, Inc., Menlo Park, CA 94025; and Departments of ^bCommunication and ^cInformation Science, Cornell University, Ithaca, NY 14853



- Effect of emotional content on people on their newsfeed on Facebook
- Negativity reduced condition - Randomly blocked posts that have negative words (e.g., sad)
- Positivity reduced condition - Randomly blocked posts that had positive words (e.g., happy)
- Negativity reduced condition - people posted more positive words and less negative words
- Positivity reduced condition - people posted more negative words and less positive words
- Participants randomly assigned to conditions without their knowledge or consent – Ethics?

Editorial Expression of Concern and Correction

PSYCHOLOGICAL AND COGNITIVE SCIENCES

PNAS is publishing an Editorial Expression of Concern regarding the following article: “Experimental evidence of massive-scale emotional contagion through social networks,” by Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock, which appeared in issue 24, June 17, 2014, of *Proc Natl Acad Sci USA* (111:8788–8790; first published June 2, 2014; 10.1073/pnas.1320040111). This paper represents an important and emerging area of social science research that needs to be approached with sensitivity and with vigilance regarding personal privacy issues.

Replace

- Replace experiments with less invasive methods, if possible

Replace

- Replace experiments with less invasive methods, if possible

Detecting Emotional Contagion in Massive Social Networks

Lorenzo Covello¹, Yunkyu Sohn², Adam D. I. Kramer³, Cameron Marlow³, Massimo Franceschetti¹, Nicholas A. Christakis^{4,5}, James H. Fowler^{2,6*}

- Detecting emotional contagion using natural experiments
- Natural experiments are where the environment creates the experiment for you
- Instead of making people sad/happy, you leverage a naturally occurring situation where people are happy/sad
- People use more negative words when it's raining, so when your friends post negative posts due to rain, how does that affect your behavior
- Not perfect — less control over the experiment — but useful in many situations

Refine

- Refine treatments to make it as less harmful as possible
- Don't want participants to suffer any more harm than is necessary
- Emotion contagion study blocked posts – criticism that people could have missed important posts because they were blocked randomly
- Alternative design - boost posts with positive/negative words
- Posts that would get knocked out would be posts that Facebook thought were less important at the bottom of the feed

Reduce

- Reduce the number of participants in your experiment as much as possible
- Emotion contagion study - used difference-of-means estimator
- Alternative design - use difference-in-difference estimator, more efficient and possible to use with smaller sample size
- Example - weight loss study with control and treatment group
 - difference-of-means - compare before and after treatment weights
 - difference-in-difference - compare the difference in weights before and after treatment
- Reducing participants even in minimal harm experiments
 - uncertainty about whether the experiment will cause harm
 - participation was not voluntary

Humane methods via 3 Rs

- With great power comes great responsibility
- Potentially more efficient than standard methods
- Opens up research in how to do research with least intrusion - e.g., differential privacy

Summary

- What are experiments, how digital experiments are different and similar to lab/field analog experiments
- Concept for designing richer experiments - validity, heterogeneity of treatment effects and mechanism
- Four ways to make digital experiments happen
- 3 Rs - Replace, Refine, Reduce

References

- <https://www.bitbybitbook.com/en/1st-ed/running-experiments/>
- <https://github.com/compsocialscience/summer-institute/tree/master/2020/materials/day6-experiments>



Questions?